76-18,403

DILLON, William Russell, 1948-
CLASSIFICATION PROBLEMS IN
MARKETING: THE CASE OF QUALITATIVE
AND/OR CATEGORICAL DATA.

City University of New York
Ph.D., 1976
Business Administration

Xerox University Microfilms, Ann Arbor, Michigan 48106

CLASSIFICATION PROBLEMS IN MARKETING: THE CASE OF QUALITATIVE AND/OR CATEGORICAL DATA

by

WILLIAM RUSSELL DILLON

A dissertation submitted to the
Graduate Faculty in Business in
partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
The City University of New York

1976

This manuscript has been read and accepted for the Graduate Faculty in Business in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 6, 1976
date

Leon Schiffman
Chairman of Examining Committee

April 6, 1976
date

Executive Officer

Professor Matthew Goldstein

Professor Edward Wolf

Dr. David Wachspress

The City University of New York

Abstract

CLASSIFICATION PROBLEMS IN MARKETING:
THE CASE OF QUALITATIVE AND/OR
CATEGORICAL DATA

by

William R. Dillon

Advisor: Professor Leon G. Schiffman

The objective of the research was to examine alternative methods of discrimination so as to provide marketing researchers with more efficient procedures for analyzing questionnaire data. In particular, the study addressed the problems associated with the use of qualitative and/or categorical data in discrimination.

The Fisher linear discriminant function has been the most frequently used classification procedure in marketing. However, in general, little attention has been given to testing the requisite assumptions for optimality and it has often been applied even when the conditions are clearly violated. In the vast majority of empirical studies using this procedure there is no mention of normality, identical variance-covariance matrices, nor any statements as to the estimation of parameters or prior probabilities. Clearly, given the discrete and often classificatory nature of marketing data, anomalies in the analysis can arise whenever techniques such as the Fisher linear discriminant

function are blindly applied. Hence, several marketing researchers have suggested that an important new direction should be the examination of discriminant procedures which can accommodate classificatory predictor variables.

Toward this end, attention was focused on two broad areas relating to the classification problem. The first stage of the study examined the relative performance of six discrimination procedures applied to binary data under a wide variety of population structures. The six discrimination procedures evaluated were: (1) full multinomial model, (2) first order independent model, (3) second order model, (4) Fisher linear discriminant model, (5) Matusita model, and (6) Martin and Bradley orthogonal polynomial model.

The second stage of the study illustrated the application of the various discrimination procedures to a data base on communication buyer behavior. The application sought to determine whether the employment of alternative discrimination procedures offer additional insights into the utilization of demographic factors in differentiating heavy and light users of a communication product.

A number of important findings evolved from examining different population structures. First, the linear models (first order and Fisher linear discriminant function) should not be used whenever the mean vectors are similar or when it is suspected, or known beforehand that the correlations are large. Second, the full multinomial, second order and Matusita procedures could better utilize the information provided by disparate correlation structures than either of the linear models. In fact, these procedures showed a marked ability to discriminate on the basis of correlation structure even when the mean differences were quite small. Third, the use of the Fisher linear discriminant function on binary data can yield severe anomalies even when the covariance matrices in the two populations are identical.

The use of the various discrimination procedures with demographic data revealed some rather interesting results. First, the full multinomial, Matusita and Martin and Bradley models yielded considerably lower error rates than the other procedures. With these procedures the error rates were all about 25 percent, whereas the use of either of the linear procedures yielded error rates about 36 percent. Second, the most critical demographic factors in differentiating heavy and light users were shown to be family income, head of household's occupation, location of previous dwelling, number of rooms in home, and length of residence. Third, a number of variables were shown to have little ability to discriminate when considered alone; however, when considered jointly with another variable their contribution to discrimination was substantially greater.

Perhaps most importantly, this research has demonstrated that extreme care should be taken in investigating the underlying structure of the data before choosing a discrimination technique. Clearly, the key to successful analysis and interpretation of marketing relationships lies in the application of procedures which better accommodate the data.

ACKNOWLEDGMENTS

First, and foremost, I wish to express my love and appreciation to my wife, Mary, who has not only quietly suffered through every phase of the dissertation process, but has lovingly sacrificed much in our three years of marriage. Without her presence, this research effort would not have been accomplished.

Likewise, my entire family is deserving of a sincere "Thank you" for providing an ideal environment in which to work. The affection and kindness they have shown will always be cherished.

The writer is grateful to Mrs. Margaret Martin and Mrs. Helen Iermiero for their careful and expedient typing. A special vote of thanks is due Mrs. Martin for making the final stage of this research effort as painless as possible.

Special thanks are extended Dr. Dan H. Moore, II, Bio-Medical Division, University of California, Lawrence Livermore Laboratory, for graciously providing several computer programs and helpful suggestions which were invaluable to the development of the final sampling methodology.

My whole-hearted thanks and appreciation are extended to my principal advisor, Professor Leon Schiffman, who not only provided help and guidance in the preparation of the dissertation, but throughout my doctoral work. His cooperation and efforts were instrumental in securing a data base, and his thoughtful comments and moral support will always be remembered.

I am also indebted to Professors Howard Gitlow and Edward Wolf for their discerning reading of the work, and their suggestions.

I take this opportunity to especially thank Professor Wolf for the time, consideration and help he has given me throughout all my years at Baruch.

Finally, a great deal of credit and gratitude must be given to Professor Matthew Goldstein, who first suggested the problem for study. I consider myself extremely fortunate to have worked with him, and my future accomplishments, if any, will reflect the values and knowledge which he has imparted. I owe him much thanks and appreciation.

TO MY WIFE,

MARY

TABLE OF CONTENTS

LIST OF TABLES

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

LIST OF FIGURES

List of Figures (continued)

CHAPTER I

INTRODUCTION

There would be little argument that marketing is a dynamic field of inquiry. Concomitant with its development there has been a gradual but pronounced shift in orientation from speculative thinking to empirical research. Such evolution requires heavy reliance on other mature disciplines for systematic inquiry generally necessitates the utilization of sophisticated behavioral and statistical methodologies. Thus, it is not surprising to note the close association between the development of behavioral constructs and the application of statistical techniques.

Even a cursory examination of the marketing literature would indicate that there has been an enormous increase in the awareness and use of complex and sophisticated procedures. The greatest proliferation seems to be in the application of multivariate techniques. The emerging emphasis on multivariate analysis is quite understandable, for the investigation of marketing phenomena is usually complex enough to render univariate or bivariate analysis inadequate.

Included under the broad umbrella of multivariate techniques is discriminant analysis. Discriminant analysis, which relates to a class of procedures, has been viewed as a kind of profile analysis, as a special form of multiple regression, or as a means of classifying individuals as belonging to one group or another more accurately than by chance.

Historically, discriminant analysis was first used as a tool in biological taxomony investigations. R.A. Fisher (1936) first addressed the general problem of allocating a new object to one of two previously established groups. His solution, which has subsequently been called the Fisher linear discriminant function (LDF) has been, for all intensive purposes, the only discrimination technique used in classification problems considered in marketing.

As is the case with many of the multivariate techniques, marketers have employed the linear discriminant model quite freely. However, the "soundness" of the LDF is dependent upon satisfying a number of rather severe assumptions regarding the parametric form of the underlying probability distribution. Although strong and restrictive, it would seem that the various constraints have not deterred marketing researchers, for they have shown a remarkable propensity to indiscriminantly use this method no matter what the structure of the data.

Surely, anomalies in the analysis will arise when techniques such as the LDF are blindly applied. This idea is further underscored since the majority of data which are analyzed in marketing studies pose some rather severe problems for they are qualitative or categorical in nature and hence the vectors (of responses) have discrete distributions. While this situation is not unique to marketing, but rather is encountered in many fields where data are derived from questionnaires, it is critical to marketing because of the questionnaire research orientation of the discipline.

Thus, it would seem crucial to examine multivariate procedures which can accommodate the type of data that are generated in the majority of marketing surveys. This is the beginning point for the following research study, the purpose of which is to investigate alternative discrimination procedures so as to provide marketing researchers with more efficient techniques for analyzing questionnaire data.

1.1 - STUDY RATIONALE

Although the Fisher linear discriminant model has been the primary statistical tool employed in numerous marketing studies (e.g., Evans, 1959; Frank and Massy, 1964; King, 1964, 1966; Massy, 1965; Robertson and Myers, 1968; Perry, 1969; Montgomery, 1975), the marketing literature lacks any substantial investigation of its compatibility with specific types of marketing data.

Frequently, either by necessity or design, the marketing researcher is faced with data the components of which are discrete valued. For example, a class of variables, generally relating to demography, are binary (classificatory) in nature. Usually, in order for demographic information to be utilized, dummy variables must be created whose values are zero or one.[1] Surely, data of this type cannot satisfy the normality constraint underlying the linear discriminant model. Thus, the common practice whereby somewhat arbitrary numeric scores are assigned to the levels or categories of the qualitative variables and then a sampled based LDF which assumes a multivariate normal structure is applied, must be viewed as being highly suspect, in addition to utilizing some rather fictitious information (e.g., measurement units are assigned to such variables as race, sex, occupation and religion).

Several researchers have been cognizant of this reality. For example, Sheth (1970), indicates that an important new direction is the extension of discriminant analysis to those situations where the predictor variables are classificatory. He comments: "This extension is quite useful in view of the fact that a large number of variables, such as sex, religion, occupation, etc., are classificatory; he continues...."It is quite surprising that Fisher suggested such extension, and still it is not widely known in marketing." (p.35)

---

[1]A perfect illustration of this can be found in the Evans (1959) study wherein 11 out of the 14 demographic were of the 0,1 type.

Another class of measurements which are frequently utilized in marketing pertain to such constructs as attitudes, opinion, cognitions and intentions. Among the most common of the measuring devices are the semantic differential, the Likert-type scale and the adjective rating scale. Generally, these take the form of seven point scales; however, it is not uncommon for responses to be dichotomous, e.g., Yes-No; Agree-Disagree; Purchase-No Purchase; or trichotomous, e.g., Low Importance-Moderately Important-Highly Important; Less Satisfaction-Equal Satisfaction-More Satisfaction. Usually, even for the case where seven point scales are used, seldom are responses distributed throughout the various states; rather the responses are clustered about a smaller number of categories, say, three or four.

Given the structure of questionnaire data, it seems highly likely that anomalies in the analysis will most assuredly occur whenever researchers blindly apply such multivariate techniques as the linear discriminant model. For the most part, a review of the literature leads to the conclusion that marketing researchers have not been predisposed to examine the compatibility of the data with the various assumptions underlying the LDF. In the vast majority of studies there is no mention of the assumption of normality nor is there any statement concerning the estimation of parameters or a priori probabilities. Generally, the assumption of equal covariances is totally neglected. In fact, a search of the literature uncovered only two empirical studies (Perry, 1969; Muezk, Mattheiss and Gable, 1974), which specifically test for homogeneity of variance.

Although the marketing literature shows a marked absence of any studies evaluating different discrimination procedures in light of data structure, a search of the statistical literature revealed several studies which provided valuable insights, and which formed the foundation for the work undertaken in this study. For example, Gilbert (1968), Revo (1970), and Moore (1970, 1973), have all been concerned with the behavior of various methods of discrimination when qualitative variables (discrete data) are used. Although Gilbert concluded that if the parameters of the distribution are "moderate" the LDF and the optimum procedure are very highly correlated, subsequent research by Revo (1970) and Moore (1970, 1973), clearly indicates that the performance of the LDF is a function of the correlation among the variables. That is, performance may degenerate when the correlation among the variables exceeds some critical value.

In particular, Moore (1973) introduces the notion of "reversals" in the log likelihood ratios (L.L.R.). This term reflects the non-monotonicity of the L.L.R.'s under certain population structures and, hence, any procedure which relies on a linear function may prove to be inadequate. To illustrate, consider the following: Let

$$X_1 = \begin{cases} 1 \text{ if individual perceives "low" risk in a new product,} \\ 0 \text{ if individual perceives "high" risk in a new product,} \end{cases}$$

and

$$X_2 = \begin{cases} 1 \text{ if individual is a "risk averter,"} \\ 0 \text{ if individual is a "risk seeker."} \end{cases}$$

Now one would expect that risk averters who perceive low risk and risk seekers who perceive high risk to be most likely to purchase the new product, i.e., should be classified into Group 1, the purchase group. On the other hand, one would expect risk seekers who perceive low risk and risk averters who perceive high risk to be least likely to purchase the new product, i.e., should be classified into Group 2, the non-purchase group. Thus, individuals with X=(0,0) and (1,1) should be classified into Group 1, while individuals with X=(1,0) and (0,1) should be classified into Group 2. If a linear procedure is employed, then denoting Z as an individual's discriminant, where

$$Z_i=b_o+b_1Z_{1i}+b_2Z_{2i}+......+b_mZ_{mi},\ i=1,2,....,n,$$

and Z* as the critical cut-off point between populations yield the following rule: Classify an individual with a score Z_i into Group 1 if $Z_i > Z^*$ and into Group 2 if otherwise. However, classifying X=(0,0) in Group 1 and X=(1,0) in Group 2 implies that $b_o > b_o + b_1$, or b_1 is negative. Similarly classifying X=(0,0) in Group 1 and X=(0,1) in Group 2 implies that b_2 is negative. Therefore, since both b_1 and b_2 are negative, then it must hold that $b_o+b_1+b_2<Z^*$,

and hence, X=(1,1) cannot be classified into the same group as X=(0,0). Misclassification with probability equal to one occurs for the L.L.R.s are not monotone.

It should be noted that Moore's study was not exhaustive for a number of severe restrictions were imposed. Monte Carlo sampling experiments were implemented which involved nineteen pairs of populations formed by specifying values for means, p_{ij}, and correlations, $r_i(j,k)$. However, only three distinct correlation patterns were considered. One group consisted of population pairs for which all correlation terms were set equal to zero. The second group contained population pairs with a single non-zero correlation, while the third group comprised those pairs in which all correlations were positive. Moore did not consider population structures in which variables had negative correlations, nor did he examine structures in which $r_1(j,k) \neq r_2(j,k)$. That is, a rather severe restriction was imposed by necessitating that the correlation between any variables in population 1 be equal to the correlation between those variables in population 2. In addition, the study utilized relatively small sample sizes, 50 and 100, in evaluating the performance of the discrimination procedures under examination. Since six variables were included in the analysis, there were 2^6, or 64 possible states. With sample sizes of 50 and 100 it was highly likely that a large number of the states were empty and therefore the stability of the estimators is somewhat suspect.

Another rationale for investigating alternative classification procedures, especially those which are of a multinomial nature, is reported by Glick (1973). In attempting to determine the convergence rates of the actual non-error to the optimum error, Glick demonstrated that for the case of qualitative variables with multinomial distributions the sample based rule for classification had actual non-error rate which converged to the optimum rate more rapidly than the multivariate normal classification rule except in cases with state j such that the true multinomial distributions place exactly equal mass, i.e., $p_j = q_j$. Here convergence rates are equivalent. In cases with state j such that $p_j \neq q_j$, the convergence of the actual non-error rate for multinomial classification procedures is exponential as the sample size increases, while the convergence speed for multivariate normal procedures is at best n^{-1}, where n is the sample size. Thus, superior multinomial convergence speed adds further credence for the investigation of, and resort to, these procedures.

Other authors (e.g., Cochran and Hopkins, 1961; Hills, 1967) have suggested the use of multinomial classification procedures, especially when the data are of a qualitative nature. However, it seems rather unlikely that marketing researchers will take this suggestion at face value for some will most assuredly argue that in the process of collapsing the data "valuable" information will be discarded. If 5-point or 7-point scales are available, then to rely on the information provided by only a dichotomy would obviously be inferior. The problems inherent in scaling are treated in Chapter III. However, suffice to say, given the available evidence on the properties of the most frequently used marketing measurement devices, the argument of information loss seems extremely weak.

It would seem that the assumptions underlying the Fisher LDF seldom hold when working with questionnaire type data, although they are often made. The ramifications of violating the assumptions discussed above are, for the most part, still open to investigation. On the one hand, it may be argued that the LDF is sound, for it is distribution free in the sense that it represents a logical procedure in that a linear combination of observations is constructed which gives the greatest amount of squared difference between the groups relative to the variance within the groups. On the other hand, the optimality of the LDF is explicitly tied to the basic assumptions of normality, equal covariances and known parameters. No matter what side of the fence one is positioned on, the evidence indicates that it is necessary (for marketing researchers) to examine alternative procedures for classification, especially given the type of data which are available in the majority of marketing studies.

1.2 - STUDY OBJECTIVES

The principal consideration in undertaking this study was to provide marketing researchers with more efficient methods for analyzing questionnaire data. Specifically, the main objectives of the study were:

1. To examine the relative performance of the following six discrimination models. (A discussion of these procedures is deferred until Chapter IV):

   Full multinomial model,
   First order-independent model,
   Second order model,
   Fisher linear discriminant model,
   Matusita model, and
   Martin and Bradley model(s).

The performance of each procedure is evaluated under a wide variety of population structures so as to report which method seems most appropriate for a particular data structure. Note, two of the above models, the Matusita as well as the Martin and Bradley, have never been examined; hence another purpose of this study is to determine whether or not these procedures merit more widespread attention.

2. To illustrate the application of these procedures on actual communication buyer behavior data, with emphasis on implementation. In particular, this stage of the study demonstrates how demographic variables can be effectively utilized as a vehicle for segmentation through the employment of alternative methods of discrimination.

This illustration seems particularly insightful especially since the prevailing consensus is that demography is only marginally important as a segmenting variable (e.g., Yankolovich, 1964; Frank, Massy, and Lonsdale, 1968; Frank, Massy, and Wind, 1972).

1.3 - QUESTIONS TO BE EXAMINED

With respect to Objective 1, which is implemented via sampling experiments, there exists an extremely large number of population structures that could be examined. Therefore, it was necessary to determine classes of population structures which seemed relevant and which facilitated the organization of the sampling so that results could be presented in an intelligible manner. Since all sampling experiments are conducted from a multinomial distribution in the absence of third and higher order terms, each discrimination procedure is evaluated in terms of the correlations among the variables. With this in mind, the first stage of the study investigates the following questions:

Question 1: Given a particular population structure, are there critical values for the correlations such that if they are exceeded the performance of certain procedures is impaired? What are the effects of negatively correlated variables on performance?

Question 2: If the difference between the mean vectors (i.e., marginal probabilities) is fixed, but the magnitude of the individual marginal probabilities are allowed to vary, what are the effects on the performance of each discrimination procedure?

Question 3: Traditionally, the discrimination problem has been addressed in terms of differences in mean structures. However, it would seem particularly interesting to determine whether certain classification methods can effectively discriminate on the basis of differences in correlation structure rather than on mean differences.

Question 4: What is the behavior of the Martin and Bradley models? How does the performance of these models compare to the other discrimination procedures?

Question 5: What are the effects of unequal sample sizes on the performance of the Matusita model? With equal sample sizes the sample based classification rule for this model can be shown to be equivalent to the usual non-parametric rule; however, the effect of unequal samples is unclear.

Although the second stage of the study principally serves as an application, the use of demographic information as a basis for describing and segmenting markets is also highlighted. Demography, recently enriched by life cycle classifications and other composite variables, has traditionally been the fundamental and most frequently used measure in segmentation studies. Nevertheless, the usefulness of demography has been questioned, and many marketing researchers have adhered to Daniel Yankelovich's (1964) plea to discard demography as a segmenting strategy in favor of more "useful" psychological criteria. It is not the purpose of this study to present an elaborate defense of demography; the task would most surely consume an entire thesis. However, the application section does seek to determine whether the employment of alternative discrimination procedures--those which are more compatible with categorical data, offer additional insights into the utilization of demographics in differentiating between groups of product users. What follows is a statement of the issues that are explored in the second stage of the study.

1. Can standard demographic variables be utilized as a vehicle for discriminating between heavy and light users of a communication product?

2. How do the results of the Monte Carlo sampling experiments of the first stage test on this data? For this data, do the error rates vary greatly among the various discrimination procedures?

3. What are the effects of dichotomizing those questions with multiple response categories? How should the variables be coded?

1.4 - SCOPE OF STUDY

This study focuses attention on two broad areas relating to the classification problem. The first stage of the study examines the performance of the various discrimination procedures under a variety of population structures, while the second stage illustrates the application of these procedures with an aim toward differentiating heavy and light users of a communication product on the basis of demographic information. The following describes the scope and limitations of this study.

1.41 - Stage 1: In this stage of the study Monte Carlo sampling experiments are initiated so as to evaluate the performance of each discrimination procedure under different population structures. A given population structure can be characterized in terms of means, p_{ij}, and correlations, $r_i(j,k)$, where the subscript i refers to a specific population. By using the Bahadur reparametization (1961), Monte Carlo samples can be generated from populations specified by the designated input parameters (p_{1j}, p_{2j}, $r_1(j,k)$, $r_2(j,k)$, and n). The parameters in this set can be varied so as to obtain different Monte Carlo samples.

The assumptions made for the purposes of sampling and evaluation are:

1. The number of populations is restricted to two and the value of Π is known; that is, the a priori probability

of group membership is fixed. In all of the Monte Carlo experiments Π is taken to 1/2.

2. The underlying distribution in the two populations is taken to be multinomial in nature.

3. The number of variables in the response vector is six. With six variables, there exists substantial savings in the number of parameters to be estimated with the second order model (21 in each population instead of the 63 required under the full multinomial). For each additional variable both time and cost for computer sampling doubles so that it becomes increasingly impractical to sample with more than six variables.

4. Each variable, X_j, is dichotomous with $p_{ij}=P(X_j=1)$ in population i. This restriction is necessary to simplify the computations and it is believed that some of the results can be extended to polychotomous X_j.

5. All of the variables are to be used in the classification scheme. Problems in obtaining a good subset of variables are considered in the second stage of the study.

6. In the majority of sampling experiments, the sample sizes from each population are equal. In many situations the researcher is able to select the sample sizes so that this restriction does not seem to be unreasonable. It was decided that sample sizes of 200 and 400

would be used in the Monte Carlo experiments in order to insure some degree of stability in the estimators. The only exception to the equal sample rule occurs for Question 4, relating to the Matusita model. For these examples, the sample size for population 1 was fixed at 200, while for population 2, the sample size was set at levels of 300 and 400, corresponding to one and half, and twice the number of observations in population 1.

Even with these restrictions, it would still require an infinite amount of time and money to sample from all possible combinations of values for the input parameters (p_{1j}, p_{2j}, $r_1(j,k)$, and $r_2(j,k)$). Therefore it was necessary to determine relevant values for all parameters. Beyond concern for generality, the only restriction placed upon admissible values was that the second order model be a probability distribution. Unfortunately, the second order model may lead to negative estimates for the cell probabilities. Hence, in all sampling experiments undertaken, values for the input parameters were chosen so that none of the estimates would be negative.

<u>1.42 - Stage II</u>: The second stage of the study considers application of the various discrimination procedures to data on communication buyer behavior. The data base was secured with the cooperation of a major United States corporation.

In order to draw the total sample from the population served by a number of associated companies, it was decided that specific subsamples would be drawn from three metropolitan areas considered representative. The areas of sampling were Atlanta, Indianapolis, and Los Angeles. The individuals selected from these areas resided in both the central city and surrounding suburban communities. In addition, an evenly-dispersed quota sample based upon respondent's estimates of their product usage rates was chosen as the method of selection.

Prior to the distribution of the questionnaire, an independent data collection firm completed a number of pretests of the questionnaire. Based upon this experience, a number of adjustments were made to the original questionnaire. The independent collection firm was also responsible for the distribution and pick-up of all questionnaires.

In distributing the questionnaire, interviewers were instructed to select a number of central city and suburban communities within each area comprising a mix of lower, middle, and upper socio-economic households. At the start of each call, interviewers screened potential respondents for the following:

1. That the household was a user of a specific communication product, and

2. That a member of the household would agree to complete the questionnaire and have the interviewer return to pick it up.

A final usable data base of 464 respondents was secured. This constituted a response rate of over 90 percent of those who agreed to complete the questionnaire.

Although this study is restricted to examining the use of demographic data as a basis for describing market segments, respondents were asked to provide information relating to three broad areas:

1. Activities, Interests, and Opinions (A.I.O.),
2. Product Usage Behavior, and
3. Demography and Socio-Economic Data.

The A.I.O. questions were of a general and product-specific nature, while the product usage section gathered information concerning type of user, usage rates, number of units in the home, and service ratings. The demography and socio-economic questions inquired into:

1. Respondent Characteristics: sex, marital status and age;
2. Head of Household Characteristics: occupation, education and age; and
3. Family Characteristics: number and age composition of children living at home, and family income.

The information gathered on family characteristics was adjusted and combined so as to provide a measure of a family's life cycle stage. The breakdown of the life cycle stages is shown below:

Code 1 Head of Household less than 55 years old, single (widowed, separated, divorced), no children;

Code 2 Head of Household less than 55 years old, married, no children;

Code 3 Head of Household less than 55 years old, with children, none teenagers;

Code 4 Head of Household less than 55 years old, with children, at least one teenager;

Code 6 Head of Household at least 55 years old, employed;

Code 7 Head of Household at least 55 years old, unemployed.

The rationale for the life cycle categories delineated above was predicated on the results and recommendations of a prior research study conducted by the company.

Although the survey was limited to only three geographical areas, it was strongly believed that these areas in total would encompass a representative cross sectional sample of product users.

1.5 - STUDY IMPORTANCE

The goal of this research study is to add to the state of knowledge with respect to the following areas:

1. This study increases the awareness of alternative methodologies for discrimination. Alternative procedures for classification are illustrated and hopefully these techniques can lead to a more efficient utilization of questionnaire data.

2. By examining alternative discrimination procedures under different classes of population structures, the marketing researcher is provided with working rules so as to determine the relevant discriminant technique. Thus, upon examination of the characteristics of the population (means and correlations), the researcher will be better able to match a discrimination procedure with the actual data, thereby reducing the number of possible anomalies.

3. This research study presents some practical statements concerning the implementation of multinomial classification procedures which seem particularly useful for discrete variables such as demography.

4. In terms of particular population structures, the results of this study should provide insights about

a. the effects of correlated independent variables on the performance of the various discrimination procedures, and

b. those procedures which are most effective when differences in mean structures are insignificant.

5. Two of the discriminant procedures used in this study (the Matusita and the Martin and Bradley model(s)) have never been investigated and, therefore, this work extends the range of knowledge in the field.

6. Perhaps most importantly, this study will provide the foundation for further research in the area of classification problems in marketing. Several areas which are not extensively addressed in this study warrant further investigation. Among these areas are: (1) the extension of the various multinomial classification procedures to polychotomous responses; (2) the development of additional procedures which can facilitate the collapsing of variables and/or states; and (3) the development of techniques which can overcome the problems associated with sparse data.

1.6 - ORGANIZATION OF THE DISSERTATION

Chapter II presents a review of the relevant marketing literature which has employed the linear discriminant model.

Chapter III is concerned with the problems associated with the LDF, with particular emphasis on the compatibility of measurement and statistical technique.

Chapter IV presents a detailed description of the various discrimination procedures together with a discussion of the criteria used to evaluate the performance of each technique. The methods used for the Monte Carlo sampling and a brief description of the computer program is also presented.

Chapter V discusses the results of the Monte Carlo sampling experiments and presents recommendations to researchers.

Chapter VI contains a description of the data base utilized in this study together with the findings and implications.

Chapter VII presents the summary and conclusions evolving from the preceding two chapters.

CHAPTER II

LITERATURE REVIEW

This chapter is devoted to a review of the pertinent marketing literature which has employed the linear discriminant model as the primary statistical vehicle for investigation. Principally, the literature review is organized around three broad areas of research studies: (1) buyer behavior phenomena; (2) diffusion of innovation; and (3) market strategy and segmentation analysis. The chapter not only reviews the major applications of discriminant analysis to marketing, but also discusses the various studies in light of the appropriateness of their statistical methodology.

2.1 - BUYER BEHAVIOR PHENOMENA

The marketing literature does not suffer from a lack of interest in the area of buyer behavior processes. The literature yields numerous studies relating to the issue of the impact of various behavioral and/or psychological variables on the buyer decision-making process. A substantial number of studies have employed the LDF with varying degrees of success. The use of discriminant analysis ranges from predicting a consumer's purchasing intentions or behavior on the basis of numerous attitudinal measures to predicting a customer's ex-post facto preferences for different types of retail institutions.

The first published marketing study employing the linear discriminant model was found in a 1951 Journal of Marketing article. In this study S. Banks investigated the relationships between preference and purchase by housewives of brands of seven classes of household products. A numerical rating scale for preferences was specifically developed for this study. The highest rating was 8- "very satisfactory," the lowest was 0- "very unsatisfactory," with the neutral point being at a scale value of 2. For two of the product categories, scouring cleanser and coffee, discriminant analysis was performed not to predict future purchase or non-purchase, but rather to evaluate the discriminatory power (relative importance) of the various preference ratings on attributes of the major brands of the two product types. Of the six product attributes measured for scouring cleanser three, cleansing ability, knowledge of price and harshness on hands, were found to have partial discriminant coefficients significantly greater than zero. For coffee, only two of the four product attributes had partial discriminant coefficients significantly different from zero-flavor and knowledge of price.

Banks does not invest much time to developing the basic ideas of discriminant analysis nor is there any mention of the assumptions underlying the LDF. He does, however, devote some attention to discussing the major differences between the linear discriminant model and regression analysis, which he also uses in the study. He indicates that the principal difference between the two methods lies in the nature of the dependent variable for regression, the dependent variable is quantitative, while the dependent variable for the discriminant is qualitative. In addition, he draws attention to the similarity between the standardized discriminant weights and the partial beta coefficients obtained in the multiple regression.

It is rather difficult to evaluate or to ascertain the appropriateness of the linear discriminant model in this study for the author, not being concerned with prediction, does not present the confusion matrix nor does he give recognition to any of the basic underlying assumptions. However, there is some evidence to indicate that measurement and statistical technique are not compatible. In discussing the collection of data, Banks suggests that before the preference data could be used for regression analysis, it had to be normalized. "The original data were piled up around 8's, 6's and 0's," (p. 150). For this reason the author thought it necessary to transform the data to ensure normality before employing regression analysis. However, he makes no comparable data transformation prior to the discriminant analysis even though the independent variables were originally identical. Skewed data is as severe a problem to the linear discriminant model as it is to regression analysis for the normality constraint is present in both techniques. Thus, certain inconsistencies exist which cast doubt on the findings of this study.

The study of brand imagery and buying behavior has received considerable attention in the literature. A classic and much examined study in this area was conducted by F.B. Evans (1959) wherein he tested the ability of psychological and objective methods to discriminate between owners of the (then) two largest selling automobiles, Ford and Chevrolet. Prior to this research, motivational researchers had painted a rather vivid portrait of buyers of these two automobile brands which indicated that these makes represented different psychological images to the public and that the purchasers of one automobile type are sharply different, psychologically speaking, from purchasers of the other. For example, Ford owners were thought to be alert to change and experiment, more tolerant, self-confident, impulsive and more masculine; while Chevrolet owners were more feminine, more cautious, suspicious, conservative, less dependent and prestige conscious.

Evans sampled from a highly restrictive and relatively homogeneous group of individuals residing in Forest Park, Illinois: Ford and Chevrolet owners of 1955-58 models; all owners were white males having only one car. Thus, with respect to psychological needs, this limited universe was believed to be highly sensitive to discrimination since there would be less confounding. The questionnaire was designed to collect three types of data: demographics, role playing questions designed to measure perceived differences and psychological needs reflecting the respondents' basic personality traits. Personal interviews were conducted to gather the first two categories of data, while the Edwards Personal Preference Schedule (EPP) was employed to measure manifest personality needs of achievement, deference, exhibition, autonomy, affiliation, intraception, dominance, abasement, change, aggression and heterosexuality.[1] A total usable data base of 140 respondents was secured: 71 Ford owners and 69 Chevrolet owners.

The results of the linear discriminant analysis indicated that an individual's personality and/or his demographic characteristics could not be effectively used to predict the choice between a Ford and Chevrolet. He

---

[1]Evans did not use the full battery of test items included in the EPP because of interviewing problems encountered in pretesting.

concluded that "the distributions of scores for all needs overlap to such an extent that discrimination is virtually impossible." (p. 368) For the personality and need variables the LDF was of doubtful statistical significance for the resulting F ratio (an analysis of variance was performed) was just barely significant at the 10 percent level; more importantly, substantial misclassification resulted. When the LDF was applied to the data (from which it was developed), it misclassified 52 individuals, or 37.1 percent of the sample (42.5 percent of the Ford owners and 31.8 percent of the Chevrolet owners). For the demographic factors, the LDF performed slightly better; it misclassified 30.1 percent of the sample. In both cases, the predictive ability leaves much to be desired for the LDF was tested under the best possible circumstances in that it was evaluated on the basis of the same sample of data from which it was developed. Thus, the stated error rates are optimistically biased.

For the most part, Evans does not examine the data in terms of assessing its compatibility with the LDF. Tests of homogeneity of variance are not performed nor is the empirical distribution of the data investigated. It should be noted that for that psychological variables (data derived by EPP) normality is somewhat more realistic for the scores of any individual could vary from 0 to 20. On the other hand, in the case of demographic and objective factors a dummy variable transformation was implemented which most assuredly produced nonnormal data. However, Evans, to his credit, does examine the correlations between personality and need variable. Intercorrelations were low and hence the need scores could be assumed to represent independent measures of personality. In addition, he was aware of the bias error rates and searches for nonlinear relationships when the linear model fails.

The results of this study have continued to interest marketers. The conclusion that psychological and objective variables are only weakly related to the ownership of two automobile brands has not been uniformly accepted and, in fact, since 1958 many comments, criticisms and reanalyses of the data have been published. Historically four types of reactions may be noted. These are:

1. Pierre Martineau (1959), disagreed with the findings, rejected the sample due to interviewer errors, and leveled a rather personal attack on Evans.

2. Gary Steiner (1961), questioned Evans' methodology and interpretation. He suggested that the design failed to consider the reliability of the criterion to be discriminated and indicated that if comparisons were made between "loyal" Ford versus "loyal" Chevrolet owners the differences would be striking. Similarly, Charles Winich (1961), suggested that a severe error was made in combining owners of the 1955-58 models. He points out that within the 1955-58 span dramatic style

changes were instituted which would make comparisons impossible. Also, Winich criticized the lumping of new car owners with those who had purchased their automobile used in addition to questioning the choice and use of the Edwards Personal Preference Schedule.

3. Ralph Westfall (1962), using a different psychological measurement device--the Thurstone Temperature Schedule--and an entirely different kind of sample, tested the same hypotheses concerning automobile brand ownership and personality types. Although not using the linear discriminant model, Westfall's findings were consistent with those of Evans; no personality differences were found between Ford and Chevrolet owners.

4. The personality data presented in the original study have been reanalyzed using different methodologies. For example, Keuhn (1963), working with only two of the need scores--dominance and affiliation--developed a method which considered the differences between these two scores. Marcus (1965), also working with the dominance and affiliation need scores, employed a graphical format in which he superimposed new axes upon the data.

Another possible limitation of the study relates to the sample sizes. By current day marketing standards, sample sizes of 71 and 69 are not considered to be sufficiently large. In addition, when attention is focused upon various subpopulations, i.e., loyal Ford owners versus loyal

Chevrolet owners, sample sizes are reduced to 29 and 39 respectively. Thus, problems peculiar to small samples may be present in the analysis. Generally speaking, the LDF performs fairly well with samples of moderate size. However, sample size is clearly a function of the parameters, number of variables and the Mahalanobis distance. It is quite possible under certain circumstances to need as many as 10 to 20 observations per variable included in the analysis (see Lachenbruch, 1975) which would necessitate larger sample sizes than those in this study.

It is interesting to note that Evans did replicate his original study. Approximately eight years after the original data were collected, a new sample was drawn from the Forest Park, Illinois, population of Ford and Chevrolet owners. In this new study, (Evans,1968), the field work, interviewing procedures, sampling, methodology, etc. were undertaken in such a manner as to closely parallel the original. A usable data base of 88 respondents, 40 Ford owners and 48 Chevrolet owners, was secured.

Paralleling the original methodology a linear discriminant function using the ten need scores was computed. Results were very similar to those obtained in the first study. As before, the ten psychological variables exhibited very little relationship to Ford or Chevrolet ownership. Thirty-four of the 88 owners, 38.6 percent, were misclassified. For the demographic and objective factors, 34.7

percent of the respondents were misclassified. Brand loyalty and imagery were also investigated and Evans implemented the methodologies suggested by Keuhn and by Marcus; however, no additional insights were derived.

Other studies have specifically examined Ford and Chevrolet owners with respect to discrimination and some have been successful. For example, Rikuma Ito (1967) successfully discriminated loyal and switching Ford and Chevrolet owners on the basis of nine attitude scales. This study specifically examines the predictive power of attitude measurements generated by the rating scale method of measuring attitude. This method requires respondents to assign an arbitrary integer among a given set of integers. The assigned integer is considered to represent the individual's attitude toward the attribute in question. In this study, respondents evaluated both brands of cars.

Ito, before employing the LDF, transforms the original attitude data into two differential measures--one representing the absolute difference between the rating given to a Ford and Chevrolet on a specific attribute; the other a relative measure derived from the ratio of the rating assigned to an attribute for both makes.

Both the absolute and relative differential measures of the attributes were found to be statistically significant, via the F test, in discriminating between Ford owners repeating Ford purchase and Ford owners switching to Chevrolet, and between Chevrolet owners repeating purchase and Chevrolet owners switching to Ford. However, with respect to prediction the results seem somewhat suspect. Ito states that the LDF's demonstrated a striking success in identifying potential switchers while leaving something to be desired in identifying potential loyal buyers. However, this statement must be discounted by the fact that total misclassification rates ranged from 36 percent to a high of 49 percent.

Further, since the LDF was evaluated on the basis of the same data from which it was developed substantial bias exists. In addition to the above, the author also does not provide any information as to what prior probabilities were used. This study employed greatly different sample sizes, only 60 out of 211 Ford owners were switchers while a very small number of Chevrolet owners indicated they had switched, 28 out of possible 366. Greatly unequal sample sizes will affect the critical discriminant score and the evaluation of the classification matrix.

Additional studies have been undertaken which attempt to discriminate between consumers' attitudes, behavior and intentions. M. Perry (1969) employed discriminant analysis to identify brand attributes that have the greatest effect on consumer behavior; that is, the LDF was applied in order to find whether an individual's purchasing intention and behavior can be predicted on the basis of his attitude toward the product. Data were obtained from a test market of dog food in two metropolitan areas. A ten point rating scale was used to measure individuals' attitudes toward the various product attributes.

With respect to the discriminant analysis, Perry does indicate that in order for the LDF to be optimal, two conditions must hold--normal distribution of variables and homogeneity of variance. Perry concedes that the data can hardly be assumed to be normally distributed but indicates that the second condition was checked and found to hold. However, he fails to note what specific test was used nor does he substantiate his statement to the effect that the second condition (homogeneity of variance) is more important than normality and therefore it (normality) will not affect the results.

Other omissions and peculiarities are present in this study. For example, there is an absence of any awareness as to the biases present when the LDF is evaluated on the basis of the same data from which it was developed. Various discriminant functions were computed with misclassification rates ranging from a low of 15 percent to a high of 49 percent. In some instances sample sizes differed greatly and a rather naive method for evaluating the confusion matrix was employed.

In the investigation of store imagery and institutional choice behavior discriminant analysis has often been employed. For example, Dodge and Summer (1969) hypothesized that discriminating patronage was related to images projected by two different types of retail institutions--specialty retailers and mass merchandisers. Based on a series of questions designed to collect data from students on (1) purchase experience, (2) institutional preferences, and (3) socio-economic status, linear discriminant analysis was applied in an attempt to classify individuals on the basis of their preference orientation toward specialty or mass merchandisers. Based upon the resultant discriminant weights, the authors conclude that experience is inversely related to shopping in a specialty-type institution. In addition, it was found that personality was negatively related to the specialty institution while the higher an individual's socio-economic level, the more he tends toward a specialty-type retailer.

The authors' use of the linear discriminant model leaves much to be desired. Although professing a desire to classify an individual's preference for one or the other type of retail institution the study never attempts to evaluate the predictive ability of the computed discriminant function on any data. In addition, the authors perform some rather questionable and unnecessary transformations on the discriminant function. For example, the original weights, $-0.00003819(x_1)$, $+ 0.00053933\ (x_2)$, $+ 0.00002395(x_3)$ were transformed so as to make the coefficient associated with x_1 have a value of 1.0000. Needless to say, the study is devoid of any remarks concerning the basic assumptions underlying the LDF.

A recent study which is closely akin to the Dodge and Summers' investigation is reported in Dash (1974). The objectives of his study were to examine and determine the impact of a group of behavioral variables on consumers' store choice decisions relating to the purchase of a specific product category. A plethora of data were collected; in fact, a total usable data base of 421 respondents was secured with 45 characteristics (independent variables) being measured.

The author attempted to discriminate between the two types of retail institutions under investigation with considerable attention being focused upon the determination of the "best" set of independent variables. Rather than applying the traditional LDF to generate a linear combination, (0,1) regression analysis was employed. This procedure, although being a theoretical sound discrimination method, does require a number of assumptions to be satisfied. Dash does not develop the distinction between the traditional LDF and the (0,1) regression approach nor does he seem aware of some of the potentially serious problems associated with this latter method. The only qualification reported (in this study) with respect to the implementation of this procedure relates to the requirement of uncorrelated independent variables. Although correct, there are, however, numerous other assumptions which must be satisfied (see, e.g., Johnston, 1972). In particular, a rather salient restriction is associated with the property of homoscadisticity. With the dependent variable only assuming a value of 0 or 1, it can be shown that the error terms (residuals) will not be normally distributed and thus the estimated mean response for an individual may be less than zero or greater than one. Theoretically, the expected value of the mean response, E(Y), for an individual must lie between zero and one, for it represents the probability that an individual possessing specific characteristics belongs to one group or the other. Again, response functions do not automatically meet this constraint. For example, in the Dash study one particular regression response function resulted in 20.5 percent of the individuals having points scores which fell outside the constraint limits (see Dash, p. 107). This phenomena is extremely problematic and can cause certain of the standard statistical tests such as the t and F tests to be inappropriate.

With respect to classification, the discriminatory procedure performed quite well. The author, to his credit, employed a split-sample validation procedure to reduce bias. The discriminant function derived from a 70 percent subsample was applied to the remaining 30 percent of the data.

In total, 74 percent of all respondents in the 30 percent sub-sample were correctly classified, while 11 percent were misclassified.[1]

Some rather novel applications of the linear discriminant model include studies conducted by P. Benson (1967) and E.H. Bonfield (1974). In the Benson study the discriminant function was used to construct scales in order to investigate whether variations in individual exposure to advertising are reflected in the individual's buying habits. In a study designed to test the Dulany theory of propositional control, Bonfield employed a discriminatory procedure to determine the relative contribution of attitude, social influence, personal norm and intention interactions as related to purchase behavior. The measurement device consisted of seven point bi-polar adjective scales; however, all data were normalized before implementing the discriminant procedure. Although principally concerned with determining the relative importance of the variables, the author did attempt to determine the predictive ability of the discriminant functions. The results indicated that the behavior models predict less well than the naive model. A split sample validation procedure was not utilized and the author correctly concludes that "some care should be therefore taken in interpreting the discriminant model." (p. 385)

---

[1] In this study, any individual possessing a point score between .4 and .6 was not classified; for this case 15 percent of the sample were not classified.

2.2 - DIFFUSION OF INNOVATION

Within the general context of social change, sociologists and anthropologists have extensively investigated the nature and substance of the diffusion process. Similarly, marketers have devoted considerable attention to the study of consumer acceptance and adoption of new products and services. In many instances, researchers have been particularly concerned with identifying those individuals who will be early adopters and, thus, the linear discriminant model has been a valuable tool.

A rather early study reported in the marketing literature was undertaken by Frank and Massy (1963). The objectives of their investigation were to determine the nature and extent of differences between families who adopted a newly introduced brand of coffee, Folger's Coffee, and those who remained with established brands. Data were collected from the Tribune's Consumer Panel from 1956 through 1958 for regular and instant coffee. In addition to demographic and socioeconomic data, purchasing records were secured for 538 families.

The primary statistical technique utilized in this study was the linear discriminant model. A total of 20 characteristics (independent variables) were amassed and three two-way discriminant functions were computed: (1) Primary and Secondary Folger's loyal versus non-Folger's households; (2) Primary loyal versus non-Folger's households; and (3) Primary versus Secondary loyal households. However, only the results for split 1 were reported in this study.

The results of the discriminant analysis reveal that, for the most part, purchasing behavior (variables) is more closely associated with household loyalty than with such socioeconomic characteristics as income and occupation. With respect to prediction, the sample LDF misclassified 25.2 percent of the respondents. It must be assumed that this error rate is derived from applying the sample LDF to the same data used to estimate the function for there is no statement to the contrary nor is the confusion matrix presented. The authors evaluate the discriminatory power of the sample LDF by comparing its error rate to that which would have been obtained if all respondents would have been classified into the larger group. This is a highly conservative procedure and may produce serious misconceptions especially if the sample sizes are greatly different. As is generally the case, the authors devote little time to discussing the linear discriminant model nor is any assessment of the compatibility of the data with the LDF.

In a study designed to examine the profile of the fashion innovator, C. King (1964) employed the linear discriminant model to differentiate the innovator (early buyer) from other consumers on the time of adoption continuum.

The data analyzed in this study were collected as part of an exploration of women's millinery. A random cluster procedure was employed, using the Metropolitan Boston Telephone Directory as the universe frame to select adult women who were then classified into adopter categories. Respondents were selected from the telephone survey and an extensive set of demographic, psychological, social, mass communication, and other relevant data were collected.

The two adopter groups analyzed were innovators or early buyers and all other consumers--late buyers and consumers that did not purchase--equal sample sizes of 74 respondents (total sample size of 148) were secured for the analysis. The 59 variables originally examined were categorized into broad sets such as Psychological Characteristics, Activity Patterns, Attitudes toward Fashion and Hats, Socioeconomic Characteristics, etc. The linear discriminant model was utilized to test the predictive power of (a) the broad sets of variables (separately) and (b) the best combination of variables.

The linear discriminant functions computed for each of the broad categories of variables proved unsatisfactory. Only one LDF was statistically significant (F test) and the misclassification rates ranged from 31.7 percent to 40.5 percent. Out of the total of 59 variables studied, the author determined a subset of 21 variables which appeared to be strong discriminators of innovator behavior (selection was on the basis of differences). Then, on the basis of these variables a sample LDF was computed. The resultant function misclassified 37 percent of the respondents.

In discussing the results of the first phase of the study, the author indicates that combinations of different types of variables are necessary in order to effectively distinguish innovators. This conclusion is based upon the rather poor discriminatory power of the various LDF's. The author, however, seems overly optimistic with respect to the results of the second phase of the analysis. The LDF based upon the 21 variables correctly classified 73 percent of the sample, and King states that this represents a 47 percent improvement over chance assignment. Although true, this is still a rather naive statement. A split sample validation procedure was not employed in the study so the stated nonerror rate is likely to be optimistically biased. In addition, it would seem impractical to base the discriminant procedure on such a large number of variables; an even smaller subset of salient variables should have been chosen for the second phase of the study. Again, there was no report of the compatibility of data with methodology and an even more severe omission resulted when the author failed to indicate the nature of the prior probabilities. King states that the data were collected via various sampling procedures; however, the reader is provided with no information as to whether the a priori probabilities were actually utilized.

The Folgers data discussed previously has supported other studies. Specifically, Frank, Massy and Morrison (1965) investigated household innovative behavior with the same Chicago Tribune Panel serving as the data base. The similarities between this study and the Frank and Massy (1963) investigation are so great that it is not necessary to discuss this work in any detail. However, it should be noted that in this study the authors realize that biases do result when the sample LDF is applied to the same data from which it was computed. The authors use a split sample technique to evaluate the predictive power of the sample LDF and comment on the magnitudes of the biases.

W. King (1963, 1966) employs the linear discriminant model to predict success or failure of a new product using early performance data from test markets. The direction of this investigation was not empirical; rather King presents a theoretical development of discriminant analysis so as to allow marketers to make predictions of eventual sales success in a newly-entered market area.

The model which is suggested is a rather novel application of the LDF. It incorporates the notion of misclassification costs and presents a formal expression for the optimal classification rule. In addition, the author discusses the estimation of parameters and reports a significant test for the discriminant function which is based upon Hotelling's extension of the t statistic to multivariate data. In comparison to previous studies, King's presentation represents a major theoretical advancement. He demonstrates that one does not have to blindly apply statistical techniques and that by understanding the basic theoretical rationale greater conceptualization may be possible.

In another study, Pessemier, Burger and Tigert (1967) employed the linear discriminant model to differentiate between triers and non-triers of a new product. The data base consisted of 265 housewives who maintained diary records for a newly introduced brand of heavy duty detergent. In total, 57 characteristics were examined; the categories consisted of such variables as: Socioeconomic Characteristics, Trial-Proneness Variables, AIO, Product Variables, etc.

The authors employed a stepwise linear discriminant analysis to differentiate between triers and non-triers of the newly introduced heavy duty detergent. Of the 57 variables, the stepwise procedure selected four variables which formed the basis for discrimination. The authors employed a split sample validation technique with good success. For the validation sample, 72 percent of the respondents were correctly classified.

Robertson and Kennedy (1968) also focus attention upon the consumer innovator. Their study utilizes the linear discriminant model to predict innovators and to assess the importance of several innovator characteristics. Data were collected from the suburban community of Deerfield, Illinois.

Innovators were defined as the first 10 percent of the community's members to adopt the small home appliance innovation which was under investigation. The sample consisted of 100 individuals, 60 innovators and 40 non-innovators.

The results of the analysis are presented in two parts. First, the authors manually compute the discriminant function assuming zero covariance among the variables. The authors state that the objective here was to "quickly" identify the important variables and to provide guidelines for the final computer analysis. The results of the step seem quite useless, for zero covariance is highly unlikely. In the second stage of the analysis the authors employed a regression methodology to produce a linear combination. Of the seven independent variables included, venturesomeness and social mobility provided the greatest contribution in discriminating between the two groups. With respect to prediction, a split-sample procedure was employed. Two validation samples were created and the misclassification rates were found to be 30.0 percent and 32 percent which were 6 and 10 percent higher than in the analysis sample.

The majority of studies undertaken in the area of diffusion of innovation have concentrated upon innovators; however, one study (Uhl, Andrus and Paulsen, 1970) has focused attention upon the last group of individuals to adopt a product, i.e., the laggards. This study attempted to determine if (1) laggards could be identified with respect to the adoption of new grocery products, and (2) distinguishing correlates of laggards could be isolated. The data base consisted of 541 households who were selected via a systematic sample of all households listed in the telephone directory for Cedar Rapids, Iowa. Of the 541 households selected, 406 completed a questionnaire designed to measure such characteristics as attitudes toward new food products, (food) venturesomeness, brand switching, or loyalty for established grocery products and various demographic factors. In total, 27 different variables were measured; however the authors selected a subset of 11 variables for the analysis. These variables were selected because they had received considerable attention in studies of innovators.

The linear discriminant model was employed to test the ability of the eleven variables to designate laggards from among all other adopters. The sample LDF misclassified 39.6 percent of the respondents. A split sample validation procedure was not utilized; thus, the resultant error rates are biased. However, the authors were aware of this problem. Although cognizant of this limitation, the authors fail to report some rather contradictory findings. In this study analysis took two forms. First, univariate tests were employed to test for significant differences among the eleven variables for the two groups. The results of these tests indicated that laggards appeared to be different only with respect to family income and brand loyalty. On the other hand, the results from the second phase of the analysis--the linear discriminant model--indicated that the most useful variable differentiating laggards was family size; the next most useful variable was life cycle, and the third most important variable was age. Family income and brand loyalty were determined to be the fourth and the sixth most useful variables respectively. Thus, the results derived from the two analyses are not compatible. It would seem that the authors are either ignorant of this fact or perceive it to be unimportant, for there is no mention of this inconsistency.

Recently, Darden and Reynolds (1974) have employed a different methodology for developing a discriminant procedure. The study attempts an investigation of male innovator behavior in the context of multiple product categories. The analysis consisted of two stages. First, respondents were clustered so as to create groups with similar ingroup innovative behavior profiles. Next the innovative behavior groups were subjected to discriminant analysis, with a large number of consumer characteristics serving as the independent variables.

The discriminant procedure suggested by Darden and Reynolds is more closely akin to principal components or factor analysis than to the traditional linear discriminant model. Here, orthogonal factors are extracted which maximize the differences among criterion group centroids. Thus, several discriminant functions (factors) are derived which best separate the groups in the discriminant space. In this study, the authors were primarily concerned with developing male innovator profiles and their methodology reflects this objective. It is quite possible that in those situations where emphasis is placed upon explanation and description rather than on prediction, this type of procedure may better accommodate the analysis.

2.3 - MARKET STRATEGY AND SEGMENTATION

Beginning with the work of W. Smith (1956), marketers have realized the benefits which can be derived from effective market segmentation. Numerous studies have suggested, recommended and utilized the linear discriminant model as a strategic management tool facilitating the development of specific marketing strategies and as a means of determining definitive market segments.

For example, W. King (1964) investigates the issues associated with performance evaluation in marketing systems and suggests that the Fisher LDF be utilized to estimate the mathematical relationships between relevant marketing performance measures and the subjective performance evaluations made by marketing executives.

Basically, King attacks the problems and complexities inherent in performance evaluations. In most instances evaluation of performance yields rather subjective adequate versus inadequate classification of some current market area. To resolve this subjectivity, the author suggests that a linear combination of performance measures be constructed which will best discriminate between those areas belonging to A (adequate) or I (inadequate). Naturally, the Fisher LDF is utilized to form the linear combination of performance measures. King develops this conceptual approach by considering an application involving the evaluation of a consumer product's performance in various market segments. In addition, two possible limitations of "descriptive" discriminant analysis are reported. The problems of bias and multicollinearity are singled out and their possible effects on the discriminant coefficients are outlined.

King has published several additional studies which have been principally concerned with this conceptual approach (e.g., King, 1966, 1967). Although illustrating the utility of this approach to different marketing environments, the various studies are, in terms of theoretical context, nearly identical. In all cases, the discriminant is employed to understand causal links between phenomena rather than for predictive purposes. The author coins the use of the LDF in this context "structural analysis."

In another study A. King and D. King (1964) utilize the linear discriminant model to determine future market shares so as to aid the firm's marketing planning process. The authors suggest that various socioeconomic, buying behavior and attitudinal characteristics be amassed for buyers and nonbuyers of a product. Thus, given observed values for these characteristics the LDF can be used to determine the likelihood that the individual will be a customer of a company or of a competitor. King and King illustrate this procedure by computing two discriminant functions. The first involves the market of the company versus the market of its primary competitors, while the second involves the market of the company versus the remaining market, excluding the primary competitor. Reported error rates were excellent--in both instances only 15 percent of the respondents were misclassified.

There are a number of problems and omissions in this study. For example, although devoting considerable time to developing the discriminant model, the authors' exposition is rather primitive. Many of the independent variables seem redundant, and it is quite possible that a great deal of multicollinearity exists. In commenting on the potential success of the model, the authors state that success or failure is dependent on: (1) establishing independent variables with high discriminatory power, and (2) ensuring that the independent variables provide useful information. However, the authors neglect any statement as to the distributional conditions which must be satisfied. The predictive power of the discriminant functions was evaluated on the basis of the same set of data from which they were derived with no apparent awareness of the biases involved. Finally, there is an absence of any recognition of the possible limitation and constraints when deriving the linear function via standard regression analysis.

W. Massy (1965) employed the linear discriminant model to differentiate market segments on the basis of audience characteristics. This study focused upon the audiences of five FM radio stations located in the Boston Metropolitan area. The data base consisted of 239 families and some 47 socioeconomic and consumption variables were measured.

The initial step in the analysis was to reduce the number of variables to a more manageable number. Factor analyzing the data produced 12 new variables which were then used in subsequent analyses. Principally, Massy utilized the LDF so as to report similarities and differences among the audiences of the five radio stations. Various discriminant functions were computed and by normalizing the confusion matrices (dividing each cell by its row total) the author was able to report those audiences which had distinctive profiles and those (audiences) which seemed to be similar to all others.

In addition to including an appendix which adequately develops the mathematics of discriminant analysis, Massy reports some rather interesting experiments. First, the number of variables included in the various discriminant functions were varied so as to illustrate the effects on discriminatory power. Secondly, the author comments on the nature and importance of the a priori probabilities. Initially, equal a priori probabilities of audience membership were assumed. Then the discriminant function was recomputed with a priori probabilities set equal to the sample frequencies. In this manner, the effects of adjusting the a priori probabilities on predicative power and interpretation were clearly shown.

In an explanatory investigation of consumer savings behavior, H. Claycamp (1965) employed the linear discriminant model to segment the thrift deposit market with respect to commercial banks and savings and loan associations. This study attempted to distinguish between customers of the two institutions on the basis of standard socioeconomic variables, personality needs (measured by EPP), motives for saving, and expectations for savings. The sample consisted of 174 randomly selected savings units who were known to hold deposits in either commercial banks (CB) or savings and loan associations (SL).

For purposes of analysis three customer groups were defined: (1) those who held thrift deposits in CB but not in SL, (2) those who held thrift deposits in SL but not in CB, and (3) those who held thrift deposits in both CB and SL. Rather than employing a three-way discriminant procedure, the author computed three two-way discriminant functions in order to evaluate the relationships among groups. The results of the discriminant analyses indicated that frequently used socioeconomic variables were of little value in discriminating between consumer groups who concentrate thrift deposits in commercial banks and those who choose savings and loan associations. In addition, the author concluded that relevant marketing strategies can best be formed on the basis of different psychological profiles found to exist between the groups. The predictive power of the various discriminant functions was evaluated; however, a split-sample validation technique was not utilized and therefore the results are inconclusive.

In a recent study, M. Etzel (1974) cites an inability to distinguish between potentially different types of credit users as a contributing factor in the unsatisfactory profit performance of bank cards. The author suggests that by utilizing the linear discriminant model to identify different types of credit card users banks can more efficiently concentrate their efforts on the most profitable segments of the market. Thus, the objectives of this study were to use demographic variables and credit related attitudes to differentiate among three types of card users: convenience users, installment users, and inactive users. The sample consisted of 1,500 randomly selected credit card holders--500 from each group.

Three board sets of characteristics were measured: attitudes toward credit in general, attitudes toward bank credit cards, and standard demographics which included age, family size, education and income. Discriminant functions were computed for each set of variables and the results were quite successful. Predictive power was good and the author was able to generate rather rich profiles of each type of user. Although not specifically examining the strong, tenuous assumptions that must be invoked, Etzel did, however, incorporate several sound precautionary procedures into the analysis. For example, all data were normalized before implementing the linear discriminant procedure and a split-sample validation was employed. In addition, the predictive power of the demographic and attitudinal variables was tested using a method suggested by Frank, Massy and Morrison (1968).

J. Muczyk, T. Mattheiss and M. Gable (1974) have adapted the linear discriminant model in order to provide a meaningful "success" profile for store managers. The cooperation of a national retail chain was secured and 34 store managers were selected for the analysis. The study design consisted of having the vice president of operations for the chain classify the 34 store managers in order of ability. Two groups were formed; the top one-half were placed in one group--high performance (HP), while the bottom one-half were placed in another--low performance (LP). All store managers were asked to complete the Ghiselli Self Dsecription Inventory--which measures seven personality characteristics.

The authors employed the LDF to identify which personality variables best characterized "successful" store managers. Of the seven personality characteristics, initiative, perceived occupational level and sociometric popularity emerged as being the most relevant. The sample LDF yielded a non-error rate of 82 percent--28 out of 34 managers were correctly classified. However, classification rates reported are suspect for the split-sample validation technique was not utilized.

In determining the "best" discriminant function the authors did not employ the usual stepwise procedure. Rather, all possible discriminant functions were examined. Where practical, this is a highly recommended procedure. Also, the authors were cognizant of the homogeneity of variance assumption for they report that the hypothesis of equal covariance matrices for the two groups could not be rejected. However, no information was provided concerning the specific test employed.

A potentially valuable and revealing study has recently been reported by D. Montgomery (1975). In addition to utilizing the traditional linear discriminant model to explore the relationship between 18 variables and a supermarket buyer's decision to accept or reject a new product the author proposes an alternative procedure for discrimination.

By examining the general appropriateness of the LDF, Montgomery was led to the area of alternative models. Specifically, the author questioned the validity of applying the linear discriminant model to vector data derived from ordinal scales, the components of which are discrete valued, and to samples for which the assumptions of normality and homogeneity of variance are obviously violated. He correctly takes the position that the strong assumptions which must be invoked suggest that the LDF should not be the only discrimination procedure examined. Thus, this study sets precedence, within the marketing literature, for the investigation of alternative methods of discrimination when the data are suspect.

Initially, Montgomery employed the traditional linear discriminant model to the sample data. The usual statistics and summary results were presented. In evaluating the confusion matrix, he notes that the discriminant function was applied to the same set of data from which it was estimated and therefore the results are optimistically biased. However, the split-sample validation technique was deemed to be impractical for the original sample sizes were rather small. He proposed an alternative validation method which involves several steps. First, the actual observations are randomly assigned to the accept or reject groups in proportion to the observed relative frequencies.

Next, the sample LDF is recomputed with the actual observations in these scrambled groups. If the discriminant results from this method are as good or nearly as good as the results obtained when the observations are in their proper groups, then spurious discriminant performance is likely. (The author suggests that several replications of this procedure be performed).

In order to resolve the obvious problems associated with data (with respect to the LDF) Montgomery proposes an alternative method of discrimination which he coins "gatekeeping analysis." This procedure involves the following steps:

1. Search for a variable and for a value of that variable which will enable us to reach a classification decision for all or a part of our sample (see step 2) while making very few errors. In effect, we seek a variable above or below which there is little or no overlap in the sample distributions from the two populations. This search may be made on the basis of prior logic and/or heuristic methods.

2. Remove from the data base those observations which we are ready to classify.

3. Return to step 1 with the remainder of the data. Repeat until sample sizes become extremely small, or no variable can be found which will achieve the objective or you are satisfied with the classification success (p. 261).

The author comments that a principal advantage of this technique lies in its applicability to relatively small samples. Although true, this method does, however, possess some rather severe limitations in that it represents an ad hoc procedure which can result in severely optimistically biased error rates.

2.4 - SUMMARY

This chapter has been devoted to a review of the relevant marketing literature which has utilized the linear discriminant model. This review has not been exhaustive, yet it attests to the widespread use of the LDF in market research studies and lends support to the notion that in some instances researchers have blindly applied this procedure.

To summarize, the following are the six basic sources of omission and misconception associated with the use of the LDF:

1. Use of data which can hardly be assumed to be compatible with the basic assumptions underlying the LDF,

2. Nonawareness of the problems associated with estimating parameters--especially the a priori probabilities,

3. Evaluating the predictive power of a sample LDF using the same data from which it was estimated,

4. A misunderstanding of the various error rates,

5. Nonawareness of the possible problems associated with multicollinear data, and

6. The blind use of (0,1) regression as an analogous procedure to the linear discriminant model.

The next chapter discusses some of the problems associated with the LDF together with the various assumptions underlying its optimality.

CHAPTER III

MEASUREMENT AND THE FISHER LINEAR DISCRIMINANT MODEL

This chapter addresses the problems associated with the proper utilization of the linear discriminant model as a vehicle for analysis. It begins with a discussion concerning the relationship between measurement and statistical technique, as well as the problems inherent with the use of marketing type data. The assumptions underlying the Fisher LDF are presented along with a discussion concerning the effect of violating these assumptions. Finally, the problem of estimating the predictive power of a discriminant function is also examined.

3.1-THE USE OF QUALITATIVE AND/OR CATEGORIC DATA

Marketers are faced with two principal dilemmas when undertaking empirical research. First, the researcher must decide upon, or develop, a means of quantifying highly subjective data representing difficult to verbalize attitudes and cognitions. (This is in addition to those problems associated with gathering "factual" information, e.g., age and income). This problem reduces to one of representing the responses of individuals to various stimuli. From a marketer's viewpoint, stimuli may evolve from consideration of alternative products, advertising copy, or package designs, and so on. The responses may involve preferences and/or adjectives which best describe the object. Regardless of the particular situation, measurement of subjective data is highly problematic for the individual's report must be taken at face value.

Secondly, the market researcher must determine the statistical technique(s) which is applicable and relevant. The investigation of marketing phenomena is complex enough to necessitate heavy reliance on statistical methodologies. The emerging emphasis on data analysis has also produced a greater reliance on "canned" computer programs. This reliance can produce severe anomalies in those situations where researchers blindly apply packaged statistical techniques without proper inspection of the underlying data structure.

These two problems, although seemingly disjointed, are, in fact, intimately related. Measurement provides the inputs which enter into statistical tables; however, the numbers that evolve from measurements have strings attached, for they carry the imprint of the operations by which they were obtained. The primacy of measurement exists for it sets the bounds on the appropriateness of statistical operations. Thus, these problems are mutually dependent and must be handled as such.

A satisfactory solution to the first problem leads one into the area of measurement and measurement scales. Unlike the physical sciences wherein measurement takes the form of well-defined scales possessing a natural zero and constant unit of measurement, researchers in the social sciences must often settle for less informative scales. For example, if a consumer is asked to rank a number of automobile brands according to overall safety, the resulting "scale" does not possess the properties associated with most physical measures.

Scales can be classified into four major categories: (1) nominal, (2) ordinal, (3) interval, and (4) ratio.[1] The importance of discussing the various scales is that each scale possesses its own characteristics or underlying assumptions regarding the matchings of elements of one domain to those of another, and the meaningfulness of performing mathematical operations on these elements.

Nominal scales are the least restrictive of the scales. Numbers which are assigned serve only as labels or tags for identifying objects, properties or events. For example, numbers are assigned to players of a game or to telephone subscribers. In the first case, each player receives a number, thus providing a one-to-one matching

---

[1]For a complete discussion of the various scales, see Green and Tull (1974) and Boyd and Westfall (1968).

of the number label and the player's name (assuming that no two or more players receive the same number label). Arithmetic operations on the numbers have no meaning with respect to the objects which they identify.

For the latter case of telephone numbers, the number-labeling process is different. For example, the first three numbers represent a particular exchange where each telephone subscriber in a particular area receives the same prefix number. Thus, all members of a specific telephone exchange are "equal" with respect to this characteristic. Again, only a very limited number of mathematical operations are allowable. It is possible to find the modal or the most numerous class and also various contingency tests may be performed. However, calculations of means, standard deviations, t tests, F tests, and so forth, are not permissible.

Ordinal scales are best characterized by rank order data. These scales require the ability to distinguish between elements in a single direction. For example, a consumer may be able to rank a number of toothpaste brands according to "flavor satisfaction." One could assign the number "1" to the highest ranking toothpaste, number "2" to the next highest ranking toothpaste, and so on. The rank order data, however, only permit comparisons of less than, greater than, or equal to; that is, rank order 2 is greater than rank order 3, and less than rank order 1. The scales do not imply that the distance between rankings is constant; the distance between rank order 2 and 1 is not necessarily equal to the distance between rank order 2 and 3. Also, with such scales there is no meaningful origin.

Ordinal scales have the characteristic of being unique up to a strictly increasing monotone transformation, i.e., one that preserves the order. Given data of this type, relatively simple statistics as the mode, median and percentiles are applicable. More sophisticated statistical measures are also applicable, such as Spearman's rank order correlation, Kendall's tau, Kendall's Q, and Goodman and Kruskal's tau. As is the case with nominal data, statistical procedures which assume equal intervals are not applicable.

Interval scales, on the other hand, not only have the same properties as ordinal scales, but also have the property of equal intervals. With interval scales, the distance between item 1 and 2 is equal to the distance between item 3 and 4. The property of equal intervals is most crucial for it allows one to use the more common (and powerful) statistical techniques such as the mean, standard deviation, and product-moment correlations, as well as the t, F, and Z tests.

Ratio scales represent the "elite" of scales in the sense that all arithmetic operations are permissible. In addition, these scales possess a meaningful origin (zero point). Equal ratios among the scales values correspond to equal ratios among the entities being measured. Ratio scales contain all the information of lower-order scales and are unique up to a positive proportionate transformation.

Of the most common measurement scales employed in marketing are the Likert-type, the semantic differential and (various forms of) the adjective ratings scale. Figure 3.1 presents examples of these three common rating scales.

For the Likert-type scale, which is perhaps the most frequently used marketing measurement device, considerable attention has been focused on the proper number of response categories. Green and Rao (1970) indicate that a 6-point or 7-point scale is optimal using reproducibility of the original data configuration as their criterion. The authors also state that this is especially true when several different instruments are employed concurrently as in a test battery. Jacoby and Matell (1971) take exception to the above findings for they do not consider data recovery as the primary criterion for most marketing research problems which employ Likert-type scales. Using reliability or validity as the principal criterion, they present evidence which indicates that a 2-point or 3-point Likert scale may be good enough. They state that "regardless of the number of steps originally employed to collect that data, conversion to dichotomous or

FIGURE 3.1

THREE POPULAR MARKETING RESEARCH SCALES

Likert-type Scale

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

Semantic Differential Scale

Good						Bad

Adjective Rating Scale

Very Good	Good	Neutral	Bad	Very Bad

Source: J. Martilla and D. Carvey. "Four Subtle Sins in Marketing Research," Journal of Marketing (January 1975), p.9.

trichotomous measures does not result in any significant decrement in reliability or validity." (p. 38) Others have examined this question, e.g., Benson (1971; Lehman and Hulbert (1972), with the general consensus being that 2-to-3 point scales are sufficient when investigating average or aggregate behavior, while 5-to-7 point scales are better suited to the investigation of individual behavior.

It would seem that several salient issues remain unresolved. For example, given that the most frequently used marketing measurement devices are often limited to 7-point scales, with dichotomous or trichotmous responses not being uncommon, can the data derived from such scales be assumed to possess properties of continuous data? Further, what are the consequences when seemingly ordinal scales are assumed to be of equal intervals, thus justifying the use of certain statistical procedures which may not be appropriate? Such questions are rather germane for it is common practice to assign numerical values to points on a scale, implicitly assuming that these numbers represent sub-equal intervals (having continuous properties) thereby allowing such simple computations as the mean and standard deviation and more sophisticated multivariate techniques as discriminant analysis to be performed.

The assumption of equal intervals is not one of semantics and should be questined on theoretical, logical and empirical grounds. Theoretically, many of the standard parametric statistical techniques assume normality. How can normality be known when only order can be ascertained? If rank order data is cubed the rank order would be the same as before; but what about normality? The assumption that a variable is normally distributed when in fact it is only amenable to ordinal measurement is a contradiction. This is somewhat magnified when discussing discriminant analysis wherein a multivariate normal distribution is assumed and where classification is based upon an <u>individual's</u> response vector (see Stevens, 1968).

Logically, there is no basis to assume that an individual will be able to judge equal units or, for that matter, that the (equal) units of one individual will correspond to those of another. Martilla and Carvey (1975) coin the use of interval scale statistics for ordinal scale data as one of the four subtle sins of marketing research. Brown and Beik (1969) argue that the psychological distance between positions 2 and 3 on an attitude scale is probably greater than the distance between positions 3 and 4. The authors state that there is "some evidence that respondents tend to avoid extremes on any judgment scale, and sometimes they avoid a neutral or zero center position, thus forming a skewed or possibly a bimodal distribution." (p. 278)

Empirically, Myers and Warner (1968) investigated adjective rating scales to determine whether the intervals are indeed subjectively equal. For their study 50 commonly used statements describing product taste or advertising effectiveness were assembled. In order to determine the consistency of meaning of the various statements from one audience to another, four separate groups of respondents were included: housewives, business executives, graduate business administration students and undergraduate business administration students. All respondents were instructed to assign the statements to a modified 21-point Thurstone differential scale.

The results of the investigation indicated that the apparent bipolar adjectives did not conform to equal intervals. For example, Table 3.1 presents a representative sample of the data obtained. The relationship suggests the subjective distance between items 1 and 2 is less than the distance between items 2 and 3 (3.02 versus 4.25). Similarly, the distance between items 4 and 5 is less than between items 3 and 4 (1.45 versus 5.89). In addition, the results indicated that some descriptive statements did not produce the expected results. "Reasonably poor" was found to be farther below the midpoint than "reasonably good" was above the midpoint, even though it would be expected that these items would be equidistant from the midpoint.

TABLE 3.1

SUBJECTIVE DISTANCE BETWEEN COMMON SCALING ADJECTIVES

Item	Scale Value	Mean Scale Value	Standard Deviation
Very good	1	16.83	2.52
Good	2	13.81	3.25
Neutral	3	9.56	1.90
Bad	4	3.67	2.54
Very bad	5	2.22	2.34

Source: J.H. Myers and G. Warner, "Semantic Properties of Selected Evaluation Adjectives," Journal of Marketing Research, Vol. 5 (November 1968), p. 41.

Besides the measurement of attitudes, opinions, cognitions and intentions, marketers often gather informative data on demography and socio-economic status. Historically, demography has been the most fundamental and frequently used measure in describing and segmenting markets; however, there would seem to be several problems with the use of demography in analysis. Generally, demographic data are of a categoric or classificatory nature as opposed to quantitative measurements. Common practice before an analysis is performed has the researcher assign numerical values to the levels of the variables or dichotomize the levels so as to produce dummy variables whose values are 0 or 1. The next step is to apply some parametric statistical procedure such as t tests, or some form of the Fisher linear discriminant function.

Surely, this type of procedure introduces anomalies in the analysis. Data of this kind can hardly be viewed as being compatible with the various assumptions underlying most of the commonly used parametric statistical tests. In addition, another problem associated with this procedure is the fictitious information introduced in the form of, say, distances between such constructs as social class or family life cycle.

A survey of the marketing literature of the last ten years indicates an enormous increase in awareness and use of complex statistical methodologies. However, what is absent is a clear recognition (by marketing researchers) of the limitations and distributional assumptions underlying the most commonly used statistical procedures. The use on ordinal scales of statistics appropriate only to interval or ratio scales must be viewed as being highly suspect. Also, since either by necessity or design marketing data are frequently of a categoric nature, anomalies are likely to arise whenever marketing researchers blindly apply statistical procedures without proper inspection of the underlying structure of the data.

3.2 - ASSUMPTIONS UNDERLYING THE FISHER LDF

The general problem of allocating a new individual based upon a battery of tests or characteristics to one of two previously established groups was apparently first studied by R.A. Fisher (1936) as a tool in biological taxonomy investigation. Under the assumption that the two groups have identical variances and covariances, and differ only in their mean responses, Fisher considered the problem of choosing the "best" linear combination of the responses. His solution was obtained by maximizing the squared difference between the groups relative to the variance within the two groups. Although the Fisher LDF is distribution free in the sense that it represents a logical procedure for constructing a linear combination, it results in an optimal (in the likelihood ratio sense) assignment rule only if a number of rather strong conditions are imposed.

What is commonly thought of as the Fisher LDF is the sample analogue of that linear function which results when applying an L.L.R. rule to two multivariate normal densities with known parameters having $\Sigma_1 = \Sigma_2 = \Sigma$, the common variance-covariance matrix. If the assumptions of normality and equal covariance matrices are satisfied, there is no need to investigate alternate classification procedures since it is well known that optimal error probabilities result when a "likelihood-ratio" rule with known parameters is used. Further, sample-based estimates of these rules are generally consistent.

However, if one or more of these conditions do not hold, the resultant linear discriminant function is not an optimal assignment rule. Given the structure of marketing data, it is highly unlikely that these conditions are in place. Data are derived from ordinal scales, are categoric in nature, and hence the vectors (of responses) have discrete distributions.

3.3 - ESTIMATION OF PREDICTIVE POWER

For the most part, marketing researchers have been cognizant of the possible bias that may arise in determining the predictive power of a particular discriminant function. Frank, Massy and Morrison (1965) discuss the bias that is produced when the sample discriminant function is applied to the same sample of data as was used to estimate the discriminant coefficients. For this case, the apparent error--simply the number of observations misclassified (same sample) is an optimistically biased estimate of the optimum error rate. Although not specifically developing the relationship between the apparent and optimum error rates, the authors do illustrate the possible magnitude of the bias. Using data which was obtained from the Chicago Tribune consumer panel, for regular coffee from 1958 through 1960, a total of 25 variables were used to develop a discriminant function relating to adopter-non-adopter categories. The total population was partitioned into smaller subsamples so as to produce five replications, and when the sample discriminant was applied to the same data an average of 72.9 percent of the observations were classified correctly. The authors indicate that the above non-error rate may be upward biased and suggest that in order to obtain a "better" measure of predictive power the sample discriminant function must be applied to an independent sample. Thus, they recommend a split-sample validation procedure whereby the data are partitioned into two samples--one sample of data is used to generate the discriminant function, while the second sample is used to ascertain the discriminatory power of the function. When applying this technique to the panel data, the average proportion classified correctly fell to

48.2 percent with a range of 44.0 to 56.0 percent across the five replications.

The authors emphasize the need for splitting the sample when estimating the predictive power of a discriminant function--and rightly so. However, the authors fail to mention several limitations of this procedure. First, although this procedure does reduce the bias, it is still not an unbiased estimate of the optimum error or the actual error. (A discussion of these error rates is deferred until Chapter IV). For the split-sample method, the error rate is given by the plug-in estimate obtained by using the estimated parameters. Hills (1966) has demonstrated empirically that the actual error rate is greater than the optimum error rate and it, in turn, is greater than the expectation of the plug-in estimate of the error rate.

In addition, this method is wasteful of data, needs large samples which are often not available, and does not evaluate the discriminant function that will be used in practice.[1] Also, the plug-in estimate of the error rate may fluctuate greatly from subsample to subsample, depending on the method of splitting the original set of observations. If the hold-out sample is large, then a good estimate of the discriminant function can be obtained

---

[1]Generally, after the discriminant function is evaluated with respect to a subsample of observations, it is recomputed using the entire sample.

but the function is likely to be poor. On the other hand, if the holdout sample is smaller, the discriminant function will be better, but the estimate of its performance will be highly variable.

3.4 - SUMMARY

This chapter has been devoted to exploring the relationship between measurement data and the Fisher linear discriminant model. The evidence presented supports the contention that care must be taken in ascertaining the compatibility of data with the assumptions underlying the LDF. This need is particularly great since with (marketing) questionnaire data, the conditions underlying the optimality of the LDF seldom hold and, hence, anomalies in the analysis may arise.

The next chapter is devoted to a development of the various discrimination procedures evaluated in this study along with a discussion of the criteria used to judge each procedure.

CHAPTER IV

PROCEDURES AND METHODOLOGY

This chapter presents a detailed description of the various discrimination procedures, together with a discussion of the criteria used to evaluate the performance of each technique. The method employed for the Monte Carlo sampling experiments and a brief description of the computer program are also contained in this chapter.

4.1 - MULTINOMIAL CLASSIFICATION

The problem of discrimination arises when there are two or more populations and it is desired to assign an individual to one of these populations on the basis of his responses to a variety of questions. Often, especially in the realm of marketing, the responses are qualitative rather than quantitative. This presents some rather interesting problems with respect to the linear discriminant model. However, there are procedures which can easily accommodate this kind of data.

A particularly useful model results when the variables X_j, $j = 1, 2, \ldots, m$, are treated as having a multinomial distribution. Let $x = (X_1, X_2, \ldots, X_m)$ be a m dimensional vector assumed to be multinomial. Suppose further that X_j takes on the values 0, 1, ..., k_j. Let $x = (x_1, \ldots, x_m)$ be called a response pattern. There are a total of

$$(k_1+1)\cdot(k_2+1)\ldots(k_m+1) = \prod_{j=1}^{m} (k_j+1) \qquad (4.1.1)$$

possible response patterns. If there are two populations called population 1 and population 2, then the discrimination problem is to assign an individual with response pattern $x=(x_1,\ldots,x_m)$ to population 1 or to population 2.

Although X_j can strictly take on any integer between 0 and k_j, for the purposes of this study a restriction is imposed in that the X_j be dichotomous with

$$\begin{aligned} p_{ij} &= P[X_j=1 \text{ in population } i], \\ 1-p_{ij} &= P[X_j=0 \text{ in population } i],\ (i=1,2). \end{aligned} \qquad (4.1.2)$$

This restriction is necessary to simplify the computations; however it would seem that some of the results can be extended to polychotomous X_j (see Appendix I). Thus, the discrimination problem can be cast in terms of m binary (Bernoulli) variables, where $x = (x_1 \ldots, x_m)$ is a vector of 0's and 1's. With m binary variables there are 2^m possible response patterns.

Now, let $\pi_i(x)$ denote the probability that response pattern $x=(x_1 \ldots, x_m)$ is observed in the i^{th} population, for $i=1,2$. Also, let Π be the probability that an individual chosen at random from the mixed population

(population 1 and population 2 combined) is in population 1. Π is called the a priori probability for population 1, and $1-\Pi$ is the a priori probability for population 2. Assuming m variables, there are $2^m=s$ possible states (response patterns) each with a corresponding $\pi_i(x)$.

The discrimination problem can be handled by determining an appropriate classification rule. A classification rule, denoted by β, can be defined such that $\beta(x)$ is equal to the probability with which a person with response pattern x is classified into population 2. Thus, $\beta(011)=1$ means that any individual with response pattern $x=(011)$ is classified into population 2 with probability equal to 1.

The Bayes classification rule which minimizes the probability that an individual chosen at random from the mixed population will be misclassified is equivalent to the following rule based upon the log likelihood ratios, $L(x)=[\pi_2(x)/\pi_1(x)]$:

$$\beta(x)=\begin{cases} 1 & \text{when } L(x) > c \\ 1-\Pi & \text{when } L(x) = c \\ 0 & \text{when } L(x) < c \end{cases} \qquad (4.1.3)$$

where $c=\log_e[\Pi/(1-\Pi)]$. Thus, the discrimination problem can be reduced to one of determining the values for c and $L(x)$, for each response pattern x. Note, for the purposes of this study, c is set to zero, since in all cases $\Pi=1-\Pi$.

Usually, the probabilities, $\pi_i(x)$, and the a priori probability, Π, are not known. However, what is available to the researcher is a random sample of individuals from each population together with their responses to the m questions (variables). Let

$n_i(x)$ = number of individuals in the sample from population i with response pattern x

and

n_i = sample size from population i.

With M populations and m variables, estimates of the a priori probability and marginal probabilities are given by

$$\Pi = n_i/n, \text{ where } n=n_1+n_2+\cdots\cdots+n_M \qquad (4.1.4)$$

and

$$p_{ij}=\sum_{S_j} n_i(x)/n_i, \text{ where } S_j=\{x:x_j=1\} \qquad (4.1.5)$$

To fix ideas, consider the following: Suppose a market researcher wishes to identify innovators of a new product based upon a number of characteristics, such as venturesomeness, perceived risk and opinion leadership. Let

X_1 = 0 if individual scores "low" on venturesomeness

1 if individual scores "high" on venturesomeness

X_2 = 0 if individual perceives "low" risk in the new product

1 if individual perceives "high" risk in the new product

X_3 = 0 if individual scores "low" on opinion leadership (nonleader)

1 if individual scores "high" on opinion leadership (leader)

Y = 0 if individual did not purchase the new product

1 if individual did purchase the new product.

With three variables there are 2^3, or eight possible distinct states:

(1) 000 (2) 001 (3) 010 (4) 011

(5) 100 (6) 101 (7) 110 (8) 111.

If the true state probabilities, $\pi_i(x)$ were not known, then they would have to be estimated. Table 4.1 illustrates the (hypothetical) results of sampling n=200 individuals from a mixture of the two populations. For this example, estimates of the a priori probabilities, and the state probabilities would be given by n_i/n, and $n_{ij}(x)/n_i$, respectively; for i= 1,2, j= 1,2....,8. That is, they are estimated on the basis of the observed frequencies. The sample-based rule is given by:

Assign x to population 2 if

$$n_{2j}(x)/n_{1j} > (1-\hat{\Pi})/1- \hat{\Pi}) \qquad (4.1.6)$$

and to population 1 if otherwise. For example,

$n_{22}(x)/n_{12}(x) = 3/13 < 1$, and therefore any individual with response pattern x = (001) would be classified as a noninnovator. A similar operation is carried out for all response patterns.

It should be mentioned that problems may arise due to zero cell frequencies. If a particular response

TABLE 4.1

HYPOTHETICAL SAMPLING RESULTS:
OBSERVED FREQUENCIES FOR n = 200

STATES (PATTERNS)

Group	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Noninnovators	3	13	52	13	3	1	13	2
Innovators	13	3	1	3	13	52	2	13

pattern is empty then it is impossible to estimate $\pi_i(x)$ for that pattern. Several procedures are available for overcoming this problem. For example, Hills (1967) suggests the use of nearest neighbor procedures. To illustrate, assume a three variable problem with possible response patterns given by

(1) 000 (2) 001 (3) 010 (4) 011

(5) 100 (6) 101 (7) 110 (8) 111

A nearest neighbor of order 1 differs from a specific pattern in only one value; a nearest neighbor of order 2 differs from a pattern in only two values, and so on. Thus, the nearest neighbors, order 1, of 000 are:

(2) 001 (3) 010 (5) 100

If the count for state j in population i is denoted by n_{ij}, then the nearest neighbor procedure assigns the observation to population 2 if

$$\frac{(n_{2j} + \sum_{(NN)_2} n_{2i})/n_2}{(n_{1j} + \sum_{(NN)_1} n_{ij})/n_1} > \hat{\Pi}/(1-\hat{\Pi}), \quad (4.1.7)$$

where $(NN)_i$ is the set of nearest neighbors to n_{ij}.

4.2 - THE PROCEDURES EVALUATED

In the last section it was shown that the problem of discriminating between two populations on the basis of m binary (Bernoulli) variables can be reduced to one of estimating 2^m log likelihood ratios given by

$$L(x)=\log_e[\,\pi_2(x)/\pi_1(x)\,] \qquad (4.2.1)$$

where $x=(x_1,\ldots.,x_m)$ is the response vector. Although the restriction that each X_j be dichotomous has been imposed, there is no restriction that the random variable X_j be uncorrelated. Thus, a variety of assumptions can be made concerning the joint distribution of the X_j in each population. Different assumptions lead to different procedures for estimating the cell probabilities $\pi_i(x)$ and the log likelihood ratios L(x). This present study considers six models for estimating $\pi_i(x)$ and L(x). For each of the following models the procedure for estimating $\pi_i(x)$ and L(x) is referenced by a bracket symbol, [*], where the asterisk inside the bracket indicates the particular model being used.

4.21 - Full Multinomial Model: Under this model no structure is assumed among the variables, X_j, of the response vector x. The model assumes the distribution of the response vector, x, in each population to be multinomial with parameters $\pi_i(x)$ in population i, for i=1, 2. The model requires the estimation of 2^m-1 independent parameters where the maximum likelihood estimators are given by

$$\pi_i(x[f]) = n_i(x)/n_i \qquad (4.21.1)$$

where $n_i(x)$ represents the number of individuals with response pattern x, in population i, and n_i is the sample size from population i.

4.22 - First Order Model: This model assumes the Bernoulli variables X_j of the response vector, x, are independent. If p_{ij} is the probability that X_j=1 in population i, then the probabilities $\pi_i(x)$ are related to the p_{ij} by

$$\pi_i(x;[1]) = \prod_{j=1}^{m} p_{ij}^{xj}(1-p_{ij})^{1-xj} \quad \text{for } i=1,2 \qquad (4.22.1)$$

Estimators of the parameters p_{ij} are given by

$$\hat{p}_{ij} = \sum_{S_j} n_i(x)/n_i \quad \text{for } i=1,\ 2 \qquad (4.22.2)$$

where S_j is the set of all patterns $(x_1,\ldots,x_m)$ with X_j=1.

This model is also called the independent model for all correlations are ignored and therefore only the marginal distribution of each X_j is used in the computation of $\pi_i(x)$. It is for this reason that a second order model which allows for correlations among the variables is included in the list of discrimination procedures.

4.23 - Second Order Model: This model assumes that all third and higher order mixed moments are zero. That is, only those correlation terms of the form $r_i(jk)$ are included. The second order model is derived from the Bahadur reparametization (See Appendix I) where the sample correlations can be expressed in the following form:

$$\hat{r}_i(jk) = [\sum_{S_{jk}} n_i(x)/n_i - \hat{p}_{ij}\hat{p}_{ik}] \cdot [\hat{p}_{ij}(1-\hat{p}_{ij})\hat{p}_{ik}(1-\hat{p}_{ik})]^{-\frac{1}{2}} \qquad (4.23.1)$$

where S_{jk} is the set of x with $X_j=1$, and $X_k=1$.
For this model the $\pi_i(x)$ are estimated by

$$\hat{\pi}_i(x;[2] = \prod_{j=1}^{m} \hat{p}_{ij}^{xj}\,(1-\hat{p}_{ij})^{1-xj} \cdot [1+\sum_{j<k} \hat{r}_i(jk)\hat{z}_{ij}\hat{z}_{ik}] \qquad (4.23.2)$$

where

$$\hat{z}_{ij} = (X_j-\hat{p}_{ij})/(\hat{p}_{ij}(1-\hat{p}_{ij}))^{\frac{1}{2}} \qquad (4.23.3)$$

This model can also be expressed in terms of the covariances between variables. This leads to approximating the probabilities, $\pi_i(x;[2])$ by

$$\prod_{j=1}^{m} \hat{p}_{ij}^{xj}(1-\hat{p}_{ij})^{1-xj} \cdot [1+\sum_{j<k} \frac{(X_j-\hat{p}_{ij})(X_k-\hat{p}_{ik})}{\hat{p}_{ij}\,\hat{q}_{ij}\,\hat{p}_{ik}\,\hat{q}_{ik}} \quad C_i(jk)], \qquad (4.23.4)$$

where $\hat{q}_{ij}=1-\hat{p}_{ij}$, and $C_i(jk)$ is the estimator for the covariance between variables X_j and X_k in population i.

Unfortunately, there is a limitation associated with the use of this model. The estimates for $\pi_i(x)$ given by (4.23.2) may lead to negative estimates. Thus, whenever this occurs the model will not lead to a probability distribution. With respect to the Monte Carlo sampling experiments described in a later section of this chapter, values for the parameters p_{ij} and $r_i(jk)$ were chosen so that none of the $\pi_i(x;[2])$ would be negative. (See Appendix I for the conditions under which the second order model leads to a probability distribution). In those cases where the sample based estimates of $\pi_i(x;[2])$ were negative, a convention whereby all negative estimates were set to 10^{-5} was evoked. This number is large enough to avoid underflow or overflow yet small enough to make the estimated log likelihood ratios large (or small) depending on whether $\pi_i(x;[2])$ occurred in the numerator (or denominator).

For the three procedures delineated above, the log likelihood ratios (L.L.R.) are simply given by

$$\hat{L}(\mathbf{x};[*])=\log_e[\pi_2(x;[*])/\pi_1(x;[*])] \qquad (4.23.5)$$

where the asterisk indicates the type of model being used.

<u>4.24 - Linear Discriminant Model</u>: The linear model results from taking the log likelihood ratio (L.L.R) of two multivariate normal probability density functions. The conditions under which the linear discriminant model results in an optimal assignment rule has been discussed in Section 3.2. It does not seem necessary to present an extensive discussion of this method; the task has already been well performed by others (e.g., Morrison, 1969; Press, 1972; Lachenbruch, 1975). However, for the purposes of this study, it would be beneficial to express the linear

discriminant function (LDF) in the case where variable X_j is Bernoulli.

For Bernoulli variables, with parameters p_{ij} in population i, the estimator for u_i, the mean, is $\hat{p}_{ij}$, and the estimator for the $j^{\underline{th}}$ row and the $k^{\underline{th}}$ column of ‡ the variance-covariance matrix is given by

$$S_{jk} = \Pi \cdot \hat{C}_1(jk) + (1-\Pi) \cdot \hat{C}_2(jk) \qquad (4.24.1)$$

where Π is the a priori probability for population 1 and $\hat{C}_i(jk)$ is the unbiased estimator of the covariance between variables X_j and X_k in population i. The LDF for correlated Bernoulli variables is given by

$$\hat{L}(x;[\ell] = \sum_j \sum_k (\hat{p}_{2j}-\hat{p}_{ij}) S^{jk} x_p$$
$$-1/2 \sum_j \sum_k (\hat{p}_{2j}-\hat{p}_{ij}) S^{jk} \cdot (\hat{p}_{2k}+\hat{p}_{1k}) \qquad (4.24.2)$$

where S^{jk} are elements of the inverse of the pooled sample covariance matrix.

With Bernoulli variables, the covariance between variables X_j and X_k may be written in terms of the correlation, r(jk), and the marginal probabilities, p_j and p_k, since

$$\text{Cov}(X_j, X_k) = r(jk) \cdot [(p_j \cdot q_j)(p_k \cdot q_k)]^{1/2}. \qquad (4.24.3)$$

Thus, it may happen that the correlation between any two variables in one population may be equal to the correlation between those same two variables in the second population; that is, $r_1(jk) = r_2(jk)$. However, if $p_{1j} \neq p_{2j}$ and/or $p_{1k} \neq p_{2k}$, the covariances will differ in the two populations.

Because of this property of binomial variables, it is possible to study cases where the covariances are equal, which is an underlying assumption used to derive the LDF, and cases where the covariances are not equal in the two populations. In general, this is not true for random variables whose distribution is multivariate normal.

4.25 - Matusita Model: This model is based upon the notion of affinity as proposed by Kameo Matusita (1952, 1955, 1956) and does not, except for regularity conditions, assume any structure among the variables.

Let F and G be distributions defined on the same space R, where $F=(p_i)$, $G=(q_i)$, $i=1,2,\ldots,k$. Then the affinity can be expressed as ρ, where

$$\rho(F,G) = \sum_{i=1}^{k}(p_i \cdot q_i)^{\frac{1}{2}} \qquad (4.25.1)$$

Letting the distance between F and G be denoted as $||F-G||$, with

$$||F-G||^2 = \sum_{i=1}^{k}(p_i^{\frac{1}{2}}-q_i^{\frac{1}{2}})^2 \qquad (4.25.2)$$

it follows that

$$||F-G||^2 = 2(1-\rho), \qquad (4.25.3)$$

or

$$||F-G|| = 2(1-\sum_{i=1}^{k}(p_i \cdot q_i)^{\frac{1}{2}}). \qquad (4.25.4)$$

Now, let S_n and S_m be the empirical distributions corresponding to F and G respectively. Therefore

$$||S_n-S_m||^2=2\left(1-\sum_{i=1}^{k}\left(\frac{n_i}{n}\cdot\frac{m_i}{m}\right)^{\frac{1}{2}}\right) \qquad (4.25.5)$$

since $S_n=(n_i/n)$, and $S_m=(m_i/m)$, where n_i and m_i are random variables representing the number of occurrences of event i, $i=1,2....,k$, in n and m observations.

A classification rule can be formulated from the above as follows. Consider classifying an obzervation Z=z into F or G using the following rule:

Classify Z into F if

$$||S_{n+1}-S_m|| \geq ||S_n-S_{m+1}||, \tag{4.25.6}$$

where $S_{n+1}(S_{m+1})$ is the empirical cumulative distribution function where z is assumed to be a sample point generated by F(G). Now, let n_i' and m_i', $i=1,2,....k$, be the state frequencies which result when z joins S_n and S_m respectively. Thus, the rule becomes

$$1-\sum_{i=1}^{k}\left[\frac{n_i'}{n+1}\cdot\frac{m_i}{m}\right]^{\frac{1}{2}} \geq 1-\sum_{i=1}^{k}\left[\frac{n_i}{n}\cdot\frac{m_i'}{m+1}\right]^{\frac{1}{2}}, \tag{4.25.7}$$

or equivalently

$$\sum_{i=1}^{k}\left[\frac{n_i}{n}\cdot\frac{m_i'}{m+1}\right]^{\frac{1}{2}} \geq \sum_{i=1}^{k}\left[\frac{n_i'}{n+1}\cdot\frac{m_i}{m}\right]^{\frac{1}{2}}. \tag{4.25.8}$$

Continuing, let the possible multinomial states be given as $E_1, E_2,....,E_k$. From F a random sample of size n is available with distribution among the classes given by $n_1, n_2,....n_k$, $\sum_{i=1}^{k} n_i=n$. Similarly, from G a random sample of size m is distributed with cell entries $m_1, m_2,....m_k$, $\sum_{i=1}^{k} m_i=m$. Suppose the observation to be classified, say Z=z, signifies state E_j, then the rule classifying Z=z into F reduces to

$$\sum_{\substack{i=1\\i\neq j}}^{k}\left[\frac{n_i}{n}\cdot\frac{m_i}{m+1}\right]^{\frac{1}{2}}+\left[\frac{n_j}{n}\cdot\frac{m_j+1}{m+1}\right]^{\frac{1}{2}}$$

$$\geq \sum_{\substack{i=1\\i\neq j}}^{k}\left[\frac{n_i}{n+1}\cdot\frac{m_i}{m}\right]^{\frac{1}{2}}+\left[\frac{n_j+1}{n+1}\cdot\frac{m_j}{m}\right]^{\frac{1}{2}} \quad . \tag{4.25.9}$$

Note, with equal sample sizes, n=m, the Matusita sample-based rule reduces to

$$(n_j(m_j+1))^{\frac{1}{2}} \geq ((n_m+1)\ m_j)^{\frac{1}{2}} \,. \tag{4.25.10}$$

This, however, is just the usual nonparametric rule:

$$\frac{n_j}{n} \geq \frac{m_j}{m} \tag{4.25.11}$$

Thus, with the restriction that n=m, the Matusita classification rule can be cast as a likelihood ratio test.

Although equivalent, this procedure does not present the usual estimation problems associated with empty cells. The fact that some of the cell frequencies may be small or even zero is not a limiting factor since the summation is extended over all k states. For the case of equal sample sizes, the log likelihood ratio is given by

$$\hat{L}(x;[M]) = \frac{\log_e((n_j+1)m_j)}{\log_e\ (n_j(m_j+1))} \quad . \tag{4.25.12}$$

For the most part, the theoretical accomplishments of Matusita have not been extended to any area of applied statistics. In fact, only recently has his notion of affinity been related to an applied field (Goldstein, Wolf and Dillon, 1976). One of the purposes of this study is to determine the usefulness of affinity (and distance) with respect to the classification problem.

4.26 - Martin and Bradley Model(s): A procedure for classifying an individual to one or two or more known populations, based upon a vector of responses, x, whose components are zero or one has recently been developed by Martin and Bradley (1972). The authors develop a model for estimating $\pi_i(x)$ which depends on a set S, of orthogonal polynomials defined on the sample space, Ω.

To illustrate, suppose there are i populations, θ_i, $i=1,2,\ldots,M$. If m characteristics are available, then the sample space Ω for x, the response vector, is discrete with 2^m sample points. Also, let the function f(x) be defined on Ω with the properties of a probability function, i.e.,

$$f(x)=\sum_{i=1}^{M} w_i\pi_i(x),\ w_i\geq 0,\ \sum_{i=1}^{M} w_i=1 \qquad (4.26.1)$$

Where $\pi_i(x)$ is a conditional probability mass function. For purposes of this study, M=2, and the mixing proportions, w_i, are set to ½. Now, a class of probability models can be defined as follows:

$$\pi_i(x) = f(x)[1+h(a^{(i)},x)] \qquad (4.26.2)$$

for $i=1,\ldots,M$; where $h(a^{(i)},x)$ is a polynomial in the elements of x and only the coefficients are specific to population i. The term $h(a^{(i)}, x)$ may be expressed in the form of orthogonal polynomials $\phi_\gamma(x)$, where

$$\phi_0=1,\quad \phi_j(x)=2X_j-1;\ j=1,\ldots .m,$$

$$\phi_\gamma(x)=\prod_{s=1}^{S}\phi_{\gamma_s}(x),\ \gamma=(\gamma_1,\gamma_2,\ldots,\gamma_s);\gamma_1<\gamma_2<\cdots<\gamma_s, \tag{4.26.3}$$

$$S=2,\ldots,m;\ \gamma_s\ \epsilon\,(1,\ldots .m).$$

The complete set of 2^m values of γ designated by Γ_m indicates all polynomial terms up to and including order m. The set of 2^m polynomials $\phi_\gamma(x)$, $\gamma\epsilon\Gamma_m$, forms an orthogonal basis for the set of all real-valued functions defined on Ω. Therefore for any $\pi_i(x)$, $i=1,\ldots,M$, $x\epsilon\Omega$, $h(a^{(i)},x)$ can be expressed as

$$h(a^{(i)},x)=\sum_{\gamma\epsilon\Gamma_m} a_\gamma^{(i)}\phi_\gamma(x). \tag{4.26.4}$$

If $f(x)\neq 0$, then

$$h(a^{(i)},x)=[\pi_i(x)-f(x)]/f(x), \tag{4.26.5}$$

for $x\epsilon\Omega$. The coefficients, $a_\gamma^{(i)}$ can be expressed as

$$a_\gamma^{(i)}=2^{-m}\sum_{x\epsilon\Omega}\phi_\gamma(x)[\pi_i(x)-f(x)]/f(x) \tag{4.26.6}$$

for $i=1,2,\ldots,m$, $\gamma\epsilon\Gamma_m$. Maximum likelihood estimators for the parameters $a_\gamma^{(i)}$ are given by

$$\hat\pi_i(x)=C^{(i)}(x)/n_i,\ i=1,2,\ldots,M,\ X\epsilon\Omega; \tag{4.26.7}$$

where $C^{(i)}(x)$ is the observed frequency for cell x for population i. It also follows that

$$\hat f(x)=\sum_{i=1}^{M} w_i\ \hat\pi_i\ (x), \tag{4.26.8}$$

and, if $a^{(i)}$ is desired, then

$$\hat a_\gamma^{(i)}=2^{-m}\sum_{x\epsilon\Omega}\phi_\gamma(x)\ Y_i(x) \tag{4.26.9}$$

for $\gamma\epsilon\Gamma_m$, where

$$Y_i(x)=[\hat{f}_i(x)-\hat{f}(x)]/\hat{f}(x) \qquad (4.26.10)$$

$i=1,2....,M$. In general, samples may be used to estimate the mixing proportions w_i in (4.26.8). However, here they are assumed to be known.

Difficulty arises when $\hat{f}(x)=0$. One of the methods for overcoming this, suggested by the authors, is to take $Y_i(x)=\sum_{\gamma\varepsilon\Gamma m}\hat{a}_\gamma\phi_\gamma(x)$ when $\hat{f}(x)>0$, but $Y_i(x) = h_I(\hat{a}_I^{(i)},x)$, $I<m$, when $\hat{f}(x)=0$. The resulting equations are then

$$D_I a_{I*}^{(i)}=C_I^{(i)} \qquad (4.26.11)$$

where the (γ,Ω)-element of D_I is

$$\sum_{x\varepsilon\delta}\Theta(x)\ \phi_\gamma(x)\phi_\delta(x),\quad \gamma,\delta\varepsilon I; \qquad (4.26.12)$$

the γ- element of $C_I^{(i)}$ is

$$\sum_{x\varepsilon s}\Theta(x)\ \phi_\gamma(x)\quad Y_i(x),\gamma\varepsilon I, \qquad (4.26.13)$$

where $\theta(x) = 1,0$ as $\hat{f}(x)\ 0=0$. The solution is in terms of $\hat{a}_{I*}^{(i)}$ which are those elements of $\hat{a}^{(i)}$ not in $\hat{a}_I^{(i)}$.

The polynomials, $\phi\gamma$, take on the values +1, and -1, and are analogous to the indepdendent variables of a 2^m factorial design in the analysis of variance model. Hence, an extremely useful aspect of this model relates to the interpretation of each $a^{(i)}$. For example, $a_o^{(i)}$ corresponds to a population characteristic analogous to a group mean; $a_j^{(i)}$ corresponds to the main effect for the j^{th} factor; $a_{j.k}^{(i)}$, $j \neq k$ corresponds to their interaction. With respect to the classification problem, $a_j^{(i)}$ measures the ability of the j^{th} dichotomous variable to discrimination; $a_{j.k}^{(i)}$ measures the joint ability of the j^{th} and k^{th} variables as discriminators, etc. Therefore, it is possible to have a situation where a variable is found not contributing to classification when considered alone, i.e., $a_j^{(i)}=0$; however, when considered jointly with another variable their contribution is found to be significant.

To fix ideas, consider the case where there are two populations and the response vector, x, is comprised of these variables. In addition, consider x=(000), and (111) as examples. For this case, $\pi_i(x)$ can be expressed in terms of f(x) and the coefficients $\hat{a}_\gamma^{(i)}$:

$$\hat{\pi}_1(000)=f(000)(1+a_o-a_1-a_2-a_3+a_{12}+a_{13}+a_{23}-a_{123}),$$

$$\hat{\pi}_2(000)=f(000)(1-a_o+a_1+a_2+a_3-a_{12}-a_{13}-a_{23}+a_{123}),$$

$$\hat{\pi}_1(111)=f(111)(1+a_o+a_1+a_2+a_3+a_{12}+a_{13}+a_{23}+a_{123}),$$

$$\hat{\pi}_2(111)=f(111)(1-a_o-a_1-a_2-a_3-a_{12}-a_{13}-a_{23}-a_{123}).$$

The above would constitute a complete model for all polynomial terms up to and including order m are included. A number of variations of the complete model can be developed. For example, the $\hat{\pi}_i(x)$ can be based solely on main effects and first order interations, i.e.,

$$\hat{\pi}_1(000)=\hat{f}(000)(1+a_o-a_1-a_2-a_3+a_{12}+a_{13}+a_{23}).$$

In addition, the model may also be approximated by using a set of lower order polynomial terms. To illustrate, take

$$h(a^{(i)},x)\approx h_s(a_s^{(i)},x)=\sum_{\gamma\epsilon\Gamma_s} a\gamma_s^{(i)}\ \phi_\gamma(x) \qquad (4.26.14)$$

with S<m; that is, S=1 or 2.

To determine a classification rule, consider the following. Let $L^{(i)}$ denote the loss associated with misclassification. These losses must be finite and may be taken to be nonnegative without loss of generality. However, it is conventional, but not necessary, to indicate a correct classification by a zero loss. Now consider classifying an individual with response vector x^*, then a minimum risk assignment rule is given by: Classify $x=x^*$ in population i if $d_i(x^*)$ is a minimum of $d\{d_1(x^*)\ldots\ldots,d_m(x^*)\}$, where

$$d_i(x)=\sum_{i=1}^{M} L^{(i)}\ \Pi(1+h(a^{(i)},x)). \qquad (4.26.15)$$

The minimum risk is given by

$$r_{\min}=\sum_{x\epsilon\Omega} f(x)\ \min_i\ (d_i(x)). \qquad (4.26.16)$$

For the case where Π is taken to be 1/2, the classification rule is simply given by: Classify $x=x^*$ in population 1 if $h(a^{(1)},x)\geq 0$; and in population 2 if otherwise.

For the purposes of this study, the following models are considered:

1. Complete Model: This model is based upon all polynomial terms up to and including order m. The complete model can be expressed as

$$\pi_i^C(x)=f(x)1+h_C(a^{(i)},x)) \qquad (4.26.17)$$

2. Incomplete Model: This model is derived from the complete model; however, only main effects and first order interactions are included in the expansion of $h(a^{(i)},x)$. Let this model be denoted by $\pi_i^I(x)$, where

$$\pi_i^I(x)=f(x)1+h_I(a_I^{(i)},x)) \qquad (4.26.18)$$

3. Reduced Model: This model results from using a set of lower order polynomials, S; i.e., S<m. Let $\pi_i^R(x)$ denote this model, where

$$\pi_i^R(x)=f(x)(1+h_R(a_R^i,x)) \qquad (4.26.19)$$

In all cases, S=2, and therefore (4.26.19) can be interpreted as a second order model.

Unfortunately, there are several limitations associated with these models. First, in dealing with data, substitution of the estimates of the parameters into each form of the model may lead to negative estimates for the probabilities. The requirement that all probabilities be positive is equivalent to

$$h(a^{(i)},x) > -1,\ f(x) > 0;\ x \varepsilon \Omega,\ i=i,2,\ldots.M, \qquad (4.26.20)$$

Whenever negative estimates occur, the value for $h(a^{(i)}, x)$ is set to -.99999. Secondly, with respect to the classification rule, the error rates obtained are apparent errors (to be defined in next section) and must be evaluated accordingly. Finally, Martin and Bradley suggest that an iterative procedure be used for estimating the coefficients, $a_{\gamma}^{(i)}$, when employing a reduced model. They state that the estimation procedure based upon the original $a_{\gamma}^{(i)}$ values may be quite poor. Since these procedures have never been evaluated, there is no evidence to indicate how poor the estimation procedure is likely to be, and hence an iterative procedure was not utilized. Thus, another purpose of this study is to report the performance of a reduced model when the original estimates for the coefficients are used.

The overwhelming advantage of these procedures lies in their ability to identify interaction effects in the context of a discrimination problem. These models not only allow the researcher to determine the contribution of each variable alone, but also the joint contribution of two or more variables can be examined. Up to the present time, there have been no reports evaluating the performance of these models. Thus, this study attempts to determine whether these procedures warrant more widespread attention.

As a point of summary, Table 4.2 presents a brief description of the advantages and disadvantages associated with each of the six discrimination procedures. For additional convenience, these models will be referred to as Full, First, Second, LDF, Matusita, Complete, Incomplete and Reduced.

4.3 - MEASUREMENT OF ERROR

Obviously, the goal of any discrimination procedure is to minimize the probability that an individual chosen at random from the mixed population is misclassified. If there are two populations, then there are two probabilities of misclassification: one for the probability that members of the first population will be misclassified, and one for the corresponding probability for members of the second population. The overall probability of misclassification, the error rate, is simply obtained by combining these two probabilities. Various authors (e.g., Cochran and Hopkins, 1961; Hills, 1966; Moore, 1970) have described a number of distinct types of error. For the purposes of this study, the following four kinds of error were computed.[1]

---

[1] A similar but more mathematical development of the behavior and properties of these types of error can be found in Moore (1970).

TABLE 4.2

COMPARISON OF DISCRIMINANT PROCEDURES

Model	Advantages	Disadvantages
Full multinomial	1. No assumptions necessary concerning structure among the variables 2. Easy to compute 3. Superior convergence	1. Too many parameters to estimate with large m 2. Possibility of many empty cells in both populations 3. No computer programs available
First order multinomial	1. Easy to compute	1. Assumption of no correlations may be too restrictive 2. Behavior with non-independent data unknown 3. No computer programs available
Second order multinomial	1. Allows for correlations 2. Easy to compute relative to LDF	1. Model can lead to negative estimates for probabilities 2. Behavior in small samples unknown 3. No computer programs available
Linear discriminant function (LDF)	1. Can use with other than 0,1 data 2. Availability of computer programs	1. Assumptions concerning underlying distributions too restrictive 2. Behavior with nonnormal data unknown

TABLE 4.2 (CONTINUED)

Model	Advantages	Disadvantages
Matusita: affinity measure	1. Easy to compute 2. Empty cells can be handled	1. Behavior in different situations unknown 2. No computer programs available
Martin and Bradley	1. Allows interactions to be examined 2. Reduced models can handle sparse data	1. Behavior is unknown 2. Error rate can only be expressed in terms of an apparent error 3. No computer programs available

4.31 - Optimum Error: The optimum error is defined as the probability of misclassification that would result if the true state probabilities $\pi_1(x)$ and $\pi_2(x)$ were known, and used to construct the classification rule. The optimum error, which will be denoted by α, can be expressed in terms of α_1 and α_2, the respective optimum probabilities that members of either population 1 or population 2 are misclassified. To illustrate: Let

$$\alpha_1 = \sum_x \beta(x) \cdot \pi_1(x) \qquad (4.31.1)$$

and

$$\alpha_2 = \sum_x [1-\beta(x)] \cdot \pi_2(x) = 1 - \sum_x \beta(x) \cdot \pi_2(x) \qquad (4.31.2)$$

Recall, β was defined in (4.1.3) by

$$\beta(x) = \begin{cases} 1 & \text{when } L(x) > c \\ 1-\Pi & \text{when } L(x) = c \\ 0 & \text{when } L(x) < c \end{cases}$$

where $L(x) = \log_e(\pi_2(x)/\pi_1(x))$ and $c = \log_e[\Pi/1-\Pi]$. Thus, the optimum error is given by

$$\alpha = \Pi \cdot \alpha_1 + (1-\Pi)\alpha_2 \qquad (4.31.3)$$

and can be computed by simply determining the log likelihood ratios from the set of state probabilities $\pi_1(x)$ and $\pi_2(x)$.

<u>4.32 - Theoretical Error</u>: When different discrimination procedures are considered another type of error can be defined. Moore (1970) labels this the "theoretical" error, and indicates that this kind of error occurs whenever a discrimination procedure other than one based on a full multinomial model is considered (p.20).

To illustrate, define the following:

$\pi_i(x;[*])$= a particular approximation of $\pi_i(x)$,

$L(x;[*])= \log_e\ (\pi_2(x;[*])\ /\ \pi_1(x;[*]))$

$L(x;[\ell])$ = the value of the LDF at x, where for the purposes of this study the asterisk again indicates the particular model being used, i.e., *=Full, First Order, Second Order or Matusita.[1] Continuing, let $\beta(x;[*])$ be the classification rule for a particular discriminant model, corresponding to [*] such that

$$\beta(x;[*] = \begin{cases} 1 & \text{if } L(x;[*]) > c \\ 1-\Pi & \text{if } L(x;[*]) = c \\ 0 & \text{if } L(x;[*]) < c \end{cases}$$

where c in all cases is taken to be zero. Now terms $\alpha_1[*]$ and $\alpha_2[*]$ can be defined by

$$\alpha_1[*]= \sum_x \beta(x;[x]).\pi_1(x)$$

and

$$\alpha_2[*]=\sum_x[1-\beta(x;[*])]\cdot\pi_2(x)$$
$$=1-\sum_x\beta(x;[*]).\pi_2(x) \qquad (4.32.1)$$

so that $\alpha_1[*]$ is equal to the probability that a member of

---

[1]For this study, the Martin and Bradley models are not evaluated in terms of a theoretical error since the sampling methodology did not allow for its computation; e.g., see Section 5.4.

population 1 is misclassified when all $\pi_i(x)$ are known and the log likelihood ratios $L(x;[*])$ are used. For members of population 2 an analogous interpretation of $\alpha_2[*]$ can be made. The theoretical error for any discrimination procedure, denoted by $\alpha[*]$ is given by

$$\alpha[*] = \Pi.\alpha_1[*]+(1-\Pi).\ \alpha_2[*] \qquad (4.32.2)$$

It is well known that unless $\beta(x) = \beta(x;[*])$ for all response patterns x, the theoretical error will be greater than the optimum error. Therefore, the optimum error is a lower bound for the theoretical error, i.e., $\alpha\leq\alpha[*]$.

<u>4.33 - Actual Error</u>: The actual error is defined as the probability of misclassification that results when the classification rule is based upon the estimated log likelihood ratios, denoted by $\hat{L}(x;[*])$. That is, it is the probability of misclassification that results from using a particular rule, $\hat{\beta}(x;[*])$. Now $A_1[*]$ and $A_2[*]$ can be defined by

$$A_1 = \sum_x \hat{\beta}(x;[*]).\pi_1(x),$$

and

$$A_2 = \sum_x (1-\hat{\beta}(x;[*]))\cdot\pi_2(x)$$
$$=1-\sum_x \hat{\beta}(x;[*]).\pi_2(x) \qquad (4.33.1)$$

where $A_i[*]$ is the probability of misclassifying a member of population i when the classification rule $\hat{\beta}(x;[*])$ is used. For the purposes of comparison, the mean (or expected) actual error is computed, which is given by

$$E\{A[*]\} = E\{\Pi . A_1[*]\} + E\{(1-\Pi) . A_2[*]\} \qquad (4.33.2)$$

Problems associated with calculating the mean actual error can be reduced by determining the distribution of the actual error through Monte Carlo sampling; this is the method utilized in this study.

Several authors have examined the relationship between the optimum error and the actual error. For example, Hills (1967), Moore (1970), and Fukunaga and Kessel (1972) have all demonstrated empirically that the optimum error is a lower bound for the actual error A[*], regardless of the procedure employed.

4.34 - Apparent Error: The apparent error is defined as the proportion of individuals from the sample used to determine the classification rule who are misclassified. It has already been noted that the use of the apparent error to evaluate a discrimination procedure can be quite misleading. Both Cochran and Hopkins (1961), and Hills (1966) have shown that on average the apparent error provides too optimistic an estimate of how well a discriminant method will perform in practice. Some indication of the possible degree of bias has been illustrated by Frank, Massy and Morrison (1965).

4.4 - EVALUATIVE CRITERIA

In order to compare the performance of one discrimination procedure against that of another, relevant criteria had to be developed. This is not an overly simple task for performance must be considered multidimensional. That is, the term performance relates to a number of measures rather than to just a single criterion.

What follows is a discussion of the performance measures used to evaluate each of the discrimination procedures.

4.41 - Increase in Theoretical Over Optimum Error: Under perfect knowledge, a researcher would obviously choose that discrimination procedure which results in the smallest possible error. Perfect knowledge for a discrimination problem would mean that the state probabilities, $\pi_i(x)$, are known and hence the optimum error could be used. However, the theoretical error (for a particular discrimination procedure) is also based on knowledge of the state probabilities and the relationship holds that $\alpha \leq \alpha[*]$. Thus, it seems particularly informative to identify those cases for which the theoretical error for a particular procedure is greater than the optimum error.

Clearly when $\alpha[k] > \alpha$, procedure k should not be recommended for even under ideal circumstances, known $\pi_i(x)$, the use of procedure k cannot result in optimal error. In other words, with perfect knowledge, a minimum requirement in choosing a particular discrimination method should be that $\alpha=\alpha[*]$.

4.42 - Mean Increase in Actual Error Over Optimum Error: The actual error is defined as the probability of misclassification that results when classification is based on the estimated log likelihood ratios. Since the expectation of the actual error is always greater than or equal to the optimum error, it is possible to define a measure of performance in terms of the difference between the mean actual error, denoted by $\bar{\Lambda}[*]$, and the optimum error, called the mean increase in actual error.

The percentage mean increase in actual error over optimum error for each discrimination procedure denoted by $\bar{\Lambda}[*]$, is given by

$$\bar{\Lambda}[*]=100.(\Lambda[*]-\alpha) \qquad (4.42.1)$$

4.43 - Mean Correlation Between Estimated L.L.R's and True L.L.R's: The mean correlation between the estimated log likelihood ratios $(\hat{L}(x;[*]))$ and the true log likelihood ratios $(L(x))$ is used as a performance measure to evaluate each of the discrimination procedures.

The mean correlation for a particular discrimination procedure, denoted by $\rho[*]$, is computed from the following formula:

$$\rho[*] = \mathrm{Cov}\{L(x),\ \hat{L}(x;[*])\} \Big/ \{\mathrm{Var}\ L(x).\mathrm{Var}\ \hat{L}(x;[*])\}^{\frac{1}{2}}. \qquad (4.43.1)$$

A correlation of 1.0 indicates perfect agreement between the estimated L.L.R.'s and the true L.L.R.'s, and hence optimal discrimination is achieved.

The criteria delineated above are by no means exhaustive for there exists other possible performance measures. For example, an alternative criterion would be that proposed by Press (1972), relating to confusion matrices. This criterion, which is derived from $Q=(N-2n)^2/N$, where N is the population size and n is the number of correct classifications, allows establishment of statistical significance since Q is distributed as χ^2 with one degree of freedom. However, since the concern here was more with comparison of the procedures rather than the establishment of statistical significance of one procedure (over chance), the four criteria outline above were judged to be more illuminating.

4.5 - PROCEDURE FOR MONTE CARLO SAMPLING

The program which carries out the Monte Carlo sampling can be summarized by the following steps:

1. The input parameters given by the set (p_{1j}, p_{2j}, $r_1(jk)$, $r_2(jk)$, and n) are used to compute the optimum error and the theoretical error for each discrimination procedure.

2. Sixty-four state probabilities (2^6), $\pi_i(x)$, are computed and these determine a set of cut-off points, which are denoted by CUT(J, I), where

$$\mathrm{CUT}(1,I) = \pi_i\ (000000)$$

$$\mathrm{CUT}(2,I) = \pi_i\ (000000) + \pi_i(000001)$$

$$\vdots$$

$$\mathrm{CUT}\ (64,I) = \pi_i(000000) + \ldots\ldots + \pi_i(111111) = 1.0 \text{ for } i=I=1,2.$$

A random number is generated, using the IBM SSP random number generator RANDU, in the interval (0,1). It is then determined between which two cut-off points the random number lies and hence a particular response pattern is identified. For example, if the random number X_o is such that

$$\mathrm{CUT}(2,I) < X_o < \mathrm{CUT}\ (3,I),$$

then $X_o=(000010)$. This process is repeated n times, where n refers to the sample sizes. In this way, two sets of observed frequencies, $n_1(x)$ and $n_2(x)$ are obtained from the multinomial distributions $\pi_1(x)$ and $\pi_2(x)$ whose underlying structure is characterized by the values of p_{ij}, p_{2j}, $r_1(jk)$ and $r_2(jk)$.

3. The observed frequencies are then used to find estimates for $\hat{p}_{1j}$, $\hat{p}_{2j}$, $\hat{r}_i(jk)$, and $C_i(jk)$, and the estimated log likelihood ratios, $\hat{L}(x;[*]$ for the various discrimination procedures. The actual errors are determined by using the rules $\beta(x;[*])$ applied to the population probabilities $\pi_i(x)$. The apparent errors for the Full, First,

Second, LDF, and Matusita models are also computed by the rules $\hat{\beta}(x;[*])$ applied to the members of the sample, while the apparent errors for the Martin and Bradley models are derived from equation (4.26.16).

4. Steps 2 and 3 are repeated 100 times for each set of input parameters. The mean actual and mean apparent errors are then computed. Also, the mean correlations between the estimated L.L.R.'s and the true L.L.R.'s are determined.

Appendix II contains a complete listing of the computer programs used in this study.

4.6 - SUMMARY

This chapter has been devoted to developing those procedures and methodologies which form the foundation for the work undertaken in this dissertation. Complete discussions of the various discrimination procedures, the four kinds of error, and the performance measures used to evaluate each discrimination procedure were presented. Finally, this chapter briefly described the Monte Carlo sampling methodology.

The next chapter presents the results for the first stage of this study--the Monte Carlo sampling experiments.

CHAPTER V

THE SAMPLING EXPERIMENTS

This chapter presents the results of the Monte Carlo sampling experiments. The sampling experiments are divided into five parts, corresponding to the five questions outlined in Section 1.3.

A sampling experiment is characterized by the values assigned to the input parameters (p_{1j}, p_{2j}, $r_1(jk)$, $r_2(jk)$, and n). Recall, the only general restriction imposed on the parameters is that they must assume values such that the second order approximation to densities is greater than or equal to zero. Although this restriction does reduce the set of admissible values, it is still impractical to sample from all possible combinations of the reduced set. Thus, in order to manage sampling, relevant ranges for means, p_{ij}, and correlations, $r_i(jk)$, were chosen.

In all cases, mean structures are characterized not only by the values assigned to the marginal probabilities, p_{1j} and p_{2j}, but also by their difference, $p_{2j}-p_{1j}$, denoted by d_p. It was decided to restrict the admissible values for the marginal probabilities in the two populations such that d_p is always less than or equal to .4. To talk to d_p

greater than .4 seems highly unrealistic for rarely are large mean differences encountered in empirical studies. Also, if large mean differences are encountered, then discrimination should easily be accomplished regardless of the particular procedure employed. That is, given large mean differences most procedures do well.

The decision as to which correlation patterns to examine proved somewhat more difficult. Here, principal concern was with generality, so as to ensure that results would be representative of population structures encountered when working with real data. Based on the kinds of correlation structures generally found in marketing studies and on other empirical research, three sufficiently general classes of correlation structures were determined. What follows is a description of the correlation structure assumed under each case.

1. Case (i): The population pairs considered under this case are assumed to have a rather simplistic correlation structure. Here, the correlation terms for population 1 are such that $r_1(jk)=0$, for all $j \neq k$. That is, the correlation matrix for population 1, denoted by R_1, is taken to be an identity matrix, $R_1=I$. On the other hand, the correlation matrix for population 2, R_2, has only one non-zero correlation term, denoted by $r_2(13)$.

2. Case (ii): For this case, the off-diagonal correlation terms in both R_1 and R_2 can take on non-zero

values. In addition, it is assumed that $R_1 \neq R_2$; i.e., $r_1(jk) \neq r_2(jk)$ for all $j \neq k$, $j,k=1,2....,6$.

3. Case (iii): An additional restriction is placed on the correlation terms for population pairs considered under this case. Here the correlation matrices in the two populations are assumed to be identical, i.e., $R_1=R_2$. In other words, the correlation between any two variables in population 1 is identical to the correlation between those two variables in population 2.

For all examples considered under Case (i) and Case (ii) results are presented in terms of the difference between correlations in the two populations, denoted by d_r, where

$$[r_2(jk)-r_1(jk)].$$

Note for Case (i), d_r is simply given by $r_2(13)$ since $R_1=I$, and $r_2(jk)=0$, for $j \neq k$, $j, k \neq 1,3$. For population pairs included under Case (iii), results are expressed in terms of r, the common value assigned to all correlation terms in both population 1 and population 2.

Before proceeding to discuss results, a note on the use of Bernoulli variables is warranted. For Bernoulli variables the covariance between variables X_j and X_k may be written in terms of the correlation, r(jk), between them and the marginal probabilities p_j and p_k since

$$Cov(X_jX_k)=r(jk)\cdot(p_jq_jp_kq_k)^{1/2},$$

where $q_j=1-p_j$. It follows that two populations have identical variance-covariance structures if and only if

$$r_1p_1q_1=r_2p_2q_2.$$

By setting $r_1(jk)\neq r_2(jk)$ the samples are taken from populations which do not necessarily have the same variance-covariance matrix. Although the assumption of equal covariances is used to derive the Fisher LDF, the marketing researcher is often faced with data, the components of which do not satisfy this constraint. Thus, there is good reason for wanting to test the ability of the LDF to discriminate when the assumption of equal covariances is violated. On the other hand, whenever $r_1(jk)=r_2(jk)$, as in Case (iii), it is possible to study examples in which the covariances are equal by simply setting $p_{2j}=1-p_{1j}$, for $j=1,2,\ldots,6$. Thus, this property of Bernoulli variables lends itself nicely to the examination of the Fisher LDF under different population structures.

The remainder of this chapter presents the results of the Monte Carlo sampling experiments.

5.1 - QUESTION 1 RESULTS

> Given a particular population structure, are there critical values for the correlations such that, if they are exceeded, the performance of certain procedures is impaired? What are the effects of negatively correlated variables on performance?

The optimum error, assuming all higher order correlations are zero, for a given set of input parameters $[p_{1j}, p_{2j}, r_1(jk), \text{ and } r_2(jk)]$ is given by

$$\alpha=\Pi.\alpha_1 + (1-\Pi).\alpha_2,$$

where α_1 and α_2 are determined by the true log likelihood ratios, $L(x)$. The optimum error is given by the full multinomial model when $\pi_1(x)$ and $\pi_2(x)$ are known. Similarly, for the second and Matusita procedures the relationship holds that

$$\pi_i(x;[2])=\pi_i(x),$$

and

$$\pi_i(x;[M])=\pi_i(x),$$

for all patterns of x and values of $[p_{1j}, p_{2j}, r_1(jk), \text{ and } r_2(jk)]$ since sampling is undertaken in the absence of third and higher order terms with equal sample sizes. However, for the first and LDF procedures there are cases where

$$\pi_i(x;[1])\neq\pi_i(x),$$

and

$$\pi_i(x;[\ell])\neq\pi_i(x).$$

Hence, it follows that the resultant theoretical errors for the second and Matusita procedures must also give optimal error for all values of the input parameters except for cases such that $n_1\neq n_2$. Here the theoretical error for the Matusita procedure does not necessarily give optimal error. For the first and LDF procedures, however, it is not generally true that $\alpha[1]=\alpha$, nor that $\alpha[\ell]=\alpha$ since there may be many patterns for which $\pi_i(x;[1]) \neq \pi_i(x)$, and $\pi_i(x;[\ell]) \neq \pi_i(x)$. It is with respect to those population structures for which $\alpha[1] \neq \alpha$ and/or $\alpha[\ell] \neq \alpha$ that interest is directed.

The initial series of Monte Carlo sampling experiments examined population pairs characterized under Case (i).

5.11 - Case (i): This case considers population pairs with a single non-zero correlation term, $r_2(13)$. In the following experiments, $r_2(13)$ takes on values -.6 to .6 in steps of .1. The values assigned to the marginal probabilities, p_{1j} and p_{2j} were fixed in a given experiment; however p_{1j} and p_{2j} were varied across experiments such that their difference, d_p, ranged between .1 and .4. In all experiments Π was set to 1/2, and each experiment consisted of 100 trials with sample sizes of 200 and 400.

The first experiment considered population pairs with $p_{1j}=.4$ and $p_{2j}=.5$, $j=1,2,....6$. Table 5.1 displays the optimum error and the theoretical errors for the first and LDF procedures as a function of d_r, where $d_r= r_2(13)$. A plot of these error rates can be found in Figure 5.1. Note neither the second nor the Matusita procedures are displayed since it has been shown that in all cases $\alpha[2]= \alpha[M]=\alpha$.

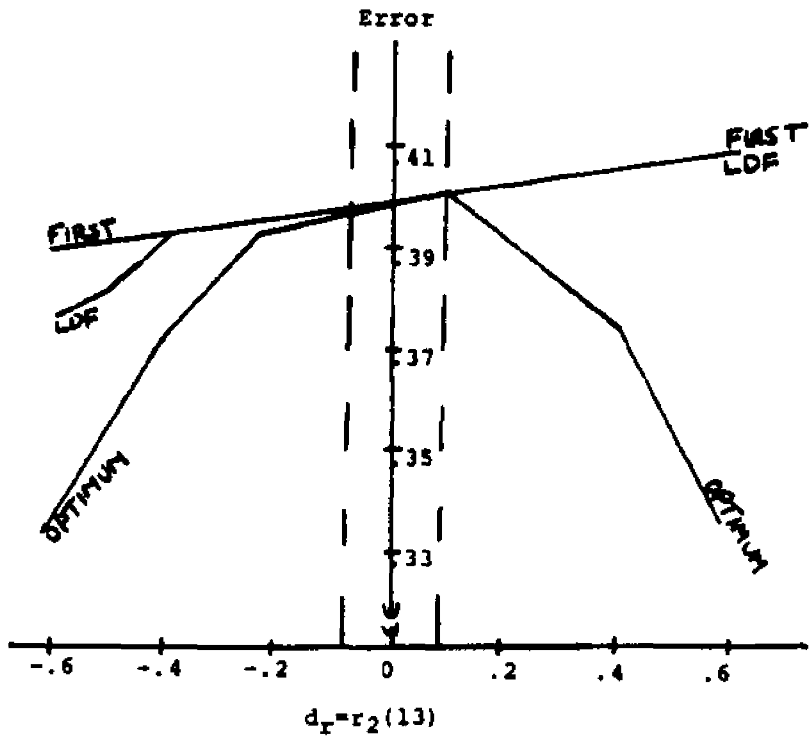

Figure 5.1 - Optimum and Theoretical Errors for $d_p = .1$ with $p_{1j} = .4$, $p_{2j} = .5$; $j = 1, 2, \ldots\ldots, 6$ and $d_r = r_2(13)$.

TABLE 5.1

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.1

WITH p_{1j}=.4, p_{2j}=.5; j=1,2,....6

d_r [r_2(13)]	Optimum Error	Theoretical First	Theoretical LDF
-.6	.3326	.3903	.3752
-.5	.3521	.3919	.3799
-.4	.3711	.3935	.3935
-.3	.3835	.3950	.3950
-.2	.3913	.3966	.3966
-.1	.3982	.3982	.3987
0	.3997	.3997	.3997
.1	.4013	.4013	.4013
.2	.3949	.4028	.4028
.3	.3871	.4044	.4044
.4	.3753	.4060	.4060
.5	.3564	.4075	.4075
.6	.3369	.4091	.4091

A study of the table and figure indicates that the theoretical errors for the first and LDF procedures can be greater than the corresponding optimum error. In fact, it is possible to define critical bounds on d_r, denoted by d_L and d_u, such that if exceeded, the use of either the first or LDF procedures result in theoretical errors which are greater than the optimum error. The vertical dashed-lines in the figure give these critical lower and upper bounds. It is apparent that if $d_r \geq -.1$ or $\leq .1$, then the theoretical errors for the first and LDF procedures are essentially the same as the optimum error.

The difference between the theoretical error and the optimum error can become quite severe, especially for $d_r > .1$ since α is a decreasing function over this range, while both $\alpha[1]$ and $\alpha[\ell]$ are monotonically increasing functions. For $d_r < -.1$, the differences are not as great, since the optimum error and the theoretical errors are monotonically decreasing over this range. Differences between the theoretical error and the optimum error for the first and LDF procedures, denoted by $\alpha[1]-\alpha$ and $\alpha[\ell]-\alpha$, are displayed in Table 5.2. A study of the table indicates that as $|d_r|$ increases, there is a point beyond which the use of either the first or LDF procedures result in substantially greater differences.

TABLE 5.2

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.1
WITH p_{1j}=.4,p_{2j}=.5;j=1,2,....,6

d_r [r_2(13)]	Optimum Error	Theoretical First $\alpha[1]-\alpha$	Theoretical LDF $\alpha[\ell]-\alpha$
-.6	.3376	5.77	4.26
-.5	.3521	3.98	2.78
-.4	.3711	2.24	2.24
-.3	.3835	1.15	1.15
-.2	.3913	.53	.53
-.1	.3982	0	0
0	.3997	0	0
.1	.4013	0	0
.2	.3949	.79	.79
.3	.3871	1.73	1.73
.4	.3753	3.07	3.07
.5	.3564	5.11	5.11
.6	.3369	7.22	7.22

The behavior of the optimum error seems particularly interesting. Although α is a decreasing function of $|d_r|$ if $|d_r| > .1$, it does not follow that $\alpha_{d_r=-.1} = \alpha_{d_r=.1}$. In fact, the optimum error is asymmetrical about d_r such that $\alpha_{-d_r} < \alpha_{d_r}$, for all values of d_r, where $-d_r$ means a $d_r < 0$. A similar relationship holds for both the first and LDF theoretical errors. Thus, performance is not only functionally related to the magnitude of the correlations, but is also dependent on the sign. Large $|d_r|$ yields better discrimination than small $|d_r|$. In other words, the evidence suggests that correlated variables may better discriminate than uncorrelated variables. This result was first reported by Elashoff, Elashoff and Goldman (1957) for two variables under more general conditions.

Attention is next focused on the mean increase in actual over optimum error which is shown in Table 5.3. Low numbers are favorable to discrimination in the table. A plot of the mean increase in actual error for each discrimination procedure can be found in Figures 5.2 and 5.3. The Matusita procedure does not appear in the figures since the sample-based classification rule for this method is equivalent to the full procedure for equal sample sizes and hence the mean increases are nearly identical. (Any differences can be attributed to sampling error). Therefore, conclusions reached for the full procedure also apply

TABLE 5.3

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR $d_p=.1$ WITH $p_{1j}=.4, p_{2j}=.5; j=1,2,\ldots,6$

d_r $[r_2(13)]$	n	Optimum Error	Full	First	Second	LDF	Matusita
	200	.3326	4.77	7.52	7.04	6.94	4.77
-.6	400		2.79	6.74	1.09	6.41	2.82
	∞		0	5.77	0	4.26	0
	200	.3521	5.01	5.59	2.44	5.25	5.01
-.5	400		3.11	4.99	1.16	4.73	3.15
	∞		0	3.98	0	2.78	0
	200	.3711	4.79	3.70	2.16	3.57	4.81
-.4	400		2.90	3.19	1.19	3.11	3.04
	∞		0	2.24	0	2.24	0
	200	.3835	4.79	2.53	2.48	2.54	4.80
-.3	400		3.22	1.98	1.32	1.86	3.29
	∞		0	1.15	0	1.15	0
	200	.3913	4.88	1.89	2.73	1.90	4.88
-.2	400		3.56	1.28	1.65	1.30	3.59
	∞		0	0.53	0	0.53	0
	200	.3982	4.71	1.33	2.71	1.40	4.71
-.1	400		3.37	0.77	1.59	0.81	3.42
	∞		0	0	0	0	0
	200	.3997	4.80	1.11	2.82	1.14	4.81
0	400		3.47	0.65	1.68	0.70	3.50
	∞		0	0	0	0	0
	200	.4013	4.62	1.12	2.66	1.25	4.61
.1	400		3.33	0.56	1.48	0.66	3.38
	∞		0	0	0	0	0
	200	.3949	4.85	1.81	2.61	1.89	4.84
.2	400		3.35	1.22	1.61	1.30	3.36
	∞		0	0.79	0	0.79	0
	200	.3871	4.75	2.60	2.35	2.72	4.76
.3	400		3.15	2.16	1.44	2.20	3.15
	∞		0	1.73	0	1.73	0
	200	.3753	4.64	3.90	2.14	4.02	4.64
.4	400		2.99	3.48	1.25	3.48	2.93
	∞		0	3.07	0	3.07	0
	200	.3564	4.82	5.84	2.14	5.70	4.80
.5	400		3.14	5.37	1.32	5.44	3.10
	∞		0	5.11	0	5.11	0
	200	.3369	5.14	7.82	2.23	7.75	5.14
.6	400		2.85	7.38	1.02	7.47	2.79
	∞		0	7.22	0	7.22	0

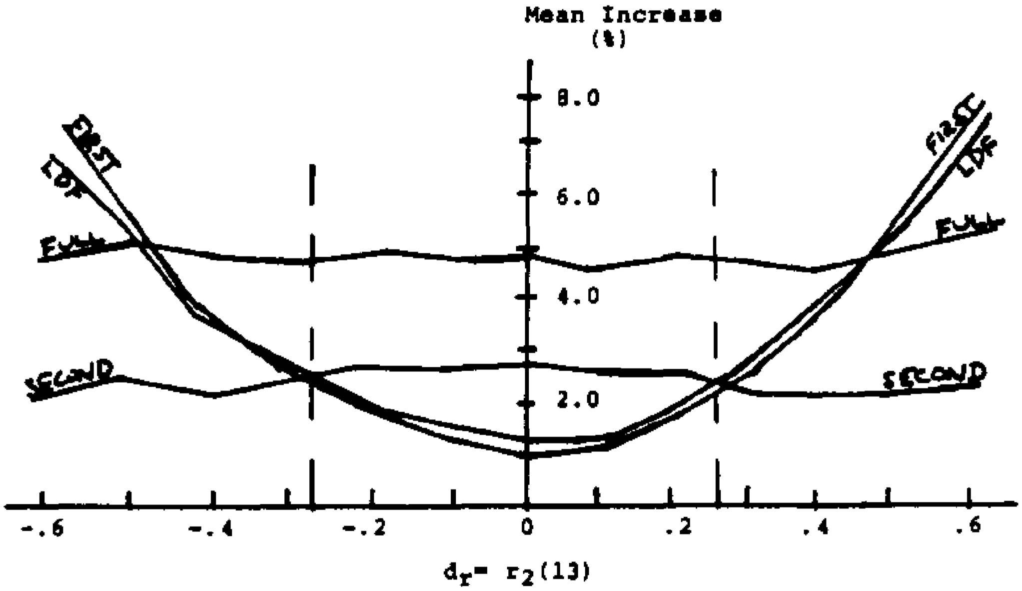

Figure 5.2 - Mean Increase in Actual Over Optimum Error (in Percent) Based on 100 Monte Carlo Trials for $d_p=.1$ with $p_{1j}=.4$, $p_{2j}=.5$; $j=1,2,\ldots\ldots,6$, $n=200$ and $d_r= r_2(13)$

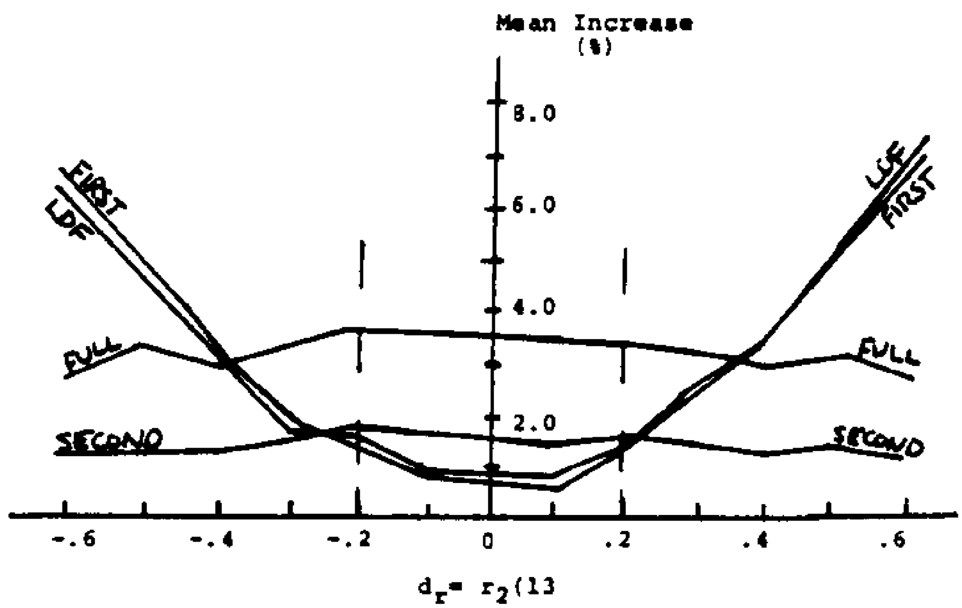

Figure 5.3 - Mean Increase in Actual Over Optimum Error (in Percent) based on 100 Monte Carlo Trials for d_p=.1 with p_{1j}=.4 p_{2j}=.5; j=1,2,....,6, n=400 and $d_r=r_2(13)$

to the Matusita model. Note the third row for each value of d_r (denoted by the infinite sample size, $n=\infty$) corresponds to the difference between the theoretical error and optimum error for each discrimination procedure.

Inspection of Figure 5.2, which is based on samples of 200, indicates that the mean increase in actual error for the full and second procedures exhibit greater stability over all values of d_r than either of the linear models. Although $\bar{\Lambda}[2] < \bar{\Lambda}[f]$ for all d_r, both of these mean increases appear relatively "flat" with respect to changes in d_r. On the other hand, the mean increase in actual error for the first and LDF procedures behave quite differently. As $|d_r|$ increases, both $\bar{\Lambda}[1]$ and $\bar{\Lambda}[\ell]$ increase and for large values of $|d_r|$, the use of either of these procedures results in significantly greater[1] mean increases. This finding is not too surprising considering that these models had differences (in theoretical over optimum error) which were increasing functions for $|d_r| > .1$.

Table 5.3 and Figure 5.3 illustrate the effects of increased sample size on the performance of each discrimination procedure. It appears that the behavior of each procedure is essentially the same across the two sets of sample sizes. However, increasing the sample size to 400

---

[1]Whenever the term "significantly greater" is used it does not imply statistical significance, since an exact test was not performed. However, this term will only be used in situations where the results for one (or more) of the models are by any criterion clearly different.

produces a general "downward shift" in each function such that if $d_p > 0$ and with fixed d_r, then

$$\bar{\Lambda}[*]_{n=400} < \bar{\Lambda}[*]_{n=200}.$$

For the most part, it would seem the effects of larger sample sizes are most pronounced on the full multinomial model.

For moderate values of d_r, the use of either the first or LDF procedures results in smaller mean increases in actual error than any other model at both n=200 and 400. However, it is apparent that large values of $|d_r|$ are quite detrimental to the performance of the linear models. With large $|d_r|$ the performance of the second order model is superior to all others. At n=200, $\bar{\Lambda}[2] < \bar{\Lambda}[*]$ for $d_r \leq -.3$ or $\geq .3$, while at n=400, a similar inequality holds for $d_r \leq -.2$ or $\geq .2$. Note for extreme values of $|d_r|$, e.g., $|d_r| \geq .5$, the full and Matusita procedures also do better than the linear models. The conclusion that the second order model out-performs all others when $|d_r|$ is large should be expected, since sampling was initiated in the absence of third and higher order terms.

Mean correlations between the observed L.L.R.'s and the true L.L.R.'s are displayed in Table 5.4. In this table large numbers are favorable to discrimination since high values of $\rho[*]$ indicate that a classification rule based on the observed L.L.R.'s agrees with the optimal classification rule (based on the true L.L.R.'s). The

TABLE 5.4

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR d_p=.1 WITH p_{1j}=.4,p_{2j}=.5; j=1,2,....6

d_r $[r_2(13)]$	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
	200	.4054	.4393	.5153	.4651	.5569	.017
-.6	400	.6386	.5528	.9065	.5358	.7035	.003
	∞	1.0	.5824	1.0	.5730	1.0	
	200	.5866	.6096	.6685	.5729	.6360	.008
-.5	400	.6159	.6621	.9128	.6586	.6899	.001
	∞	1.0	.6709	1.0	.6637	1.0	
	200	.4870	.6963	.5729	.6730	.5676	.003
-.4	400	.7211	.7478	.9255	.7457	.7242	-
	∞	1.0	.7608	1.0	.7561	1.0	
	200	.3213	.6695	.7768	.7107	.4195	.002
-.3	400	.7167	.8082	.8242	.8114	.7233	-
	∞	1.0	.8487	1.0	.8459	1.0	
	200	.3307	.8412	.6263	.8691	.4382	.001
-.2	400	.6080	.8875	.8070	.8980	.6533	-
	∞	1.0	.9258	1.0	.9246	1.0	
	200	.2874	.8919	.5986	.9045	.3544	.001
-.1	400	.3492	.8761	.7977	.8711	.4087	-
	∞	1.0	.9802	1.0	.9799	1.0	
	200	.3114	.7713	.5868	.7783	.3793	.001
0	400	.5238	.9638	.8859	.9545	.6237	-
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.4272	.9443	.5304	.9570	.4815	.001
.1	400	.4254	.9058	.8336	.8935	.5033	-
	∞	1.0	.9802	1.0	.9799	1.0	
	200	.2811	.8903	.7759	.8755	.4380	-
.2	400	.5623	.8677	.8748	.8736	.5999	-
	∞	1.0	.9258	1.0	.9249	1.0	
	200	.4622	.6045	.7149	.6332	.4522	.002
.3	400	.6821	.7937	.8164	.7786	.6850	-
	∞	1.0	.8487	1.0	.8468	1.0	
	200	.3518	.7222	.7798	.7239	.4435	.002
.4	400	.7264	.7417	.9148	.7290	.7330	-
	∞	1.0	.7608	1.0	.7581	1.0	
	200	.5068	.6317	.6878	.6038	.5694	.007
.5	400	.6076	.6531	.8678	.6451	.7018	.001
	∞	1.0	.6706	1.0	.6681	1.0	
	200	.5669	.4579	.5744	.4366	.6468	.016
.6	400	.5936	.5600	.9258	.5617	.7284	.001
	∞	1.0	.5834	1.0	.5783	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

last column of the table shows the average proportion of times the second order Bahadur approximation to densities gave negative results.

The first point to be made from the table relates to the third row for each value of d_r (denoted in the table by the infinite sample size, $n=\infty$). This row gives the value of the performance criterion when the true state probabilities, $\pi_i(x)$, are known. An examination of these rows indicates that there is only one value for d_r such that all $\rho[*]$ are equal to 1.0. (Ideally, a correlation of 1.0 is desired since it indicates perfect agreement). For $d_r \neq 0$, the mean correlations for both the first and LDF procedures are less than 1.0. Hence, the L.L.R.'s for these procedures are linearly different from the true L.L.R.'s even when the true state probabilities are known.

To talk to general trends in the various $\rho[*]$ as d_r varies presents some difficulty, for the mean correlations are non-monotonic. This is not surprising for the L.L.R.'s themselves exhibit a good deal of non-monotonicity. Nevertheless, the table still provides information as to the superiority of one procedure over another for fixed d_r. For example, at $n=400$, the second procedure yields higher mean correlations than the other procedures if $d_r < -.3$ or $> .2$. For d_r between $-.2$ and $.1$, the use of either the first or LDF results in significantly higher mean correlations than the full or Matusita procedures. Also, the performance of the

Matusita model is somewhat better than the full procedure for all values of d_r; this is especially true at n=200. In part, this may be due to the way in which the Matusita model handles empty cells. For extreme d_r, the performance of the linear models again falls off and the use of these models results in mean correlations which are significantly lower than the rest.

The average proportion of times the second order approximation $\pi_i(x;[2])$ was negative appears in the last column of the table. It would seem that the estimation procedure was satisfactory since the proportions appearing in this column are rather small. At n=400, nine out of the 13 population structures had all $\pi_i(x;[2]) \geq 0$, while at n=200, the largest average proportion appears for d_r = 6. and even for this structure only 1.7 percent of the estimates were negative.

In order to test whether the conclusions delineated above would hold for other values, further sampling was initiated for d_p=.2, .3, and .4 with corresponding marginal probability vectors given by

$$p_{1j}=(.2,.2,.2,.2,.2,.2) \quad ; \quad p_{2j}=(.4,.4,.4,.4,.4,.4),$$
$$p_{1j}=(.3,.3,.3,.3,.3,.3) \quad ; \quad p_{2j}=(.6,.6,.6,.6,.6,.6),$$
$$p_{1j}=(.2,.2,.2,.2,.2,.2) \quad ; \quad p_{2j}=(.6,.6,.6,.6,.6,.6).$$

Also of interest was how larger mean differences would affect the various performance measures. Summary results for the optimum error and theoretical errors are presented in

Tables 5.5 through 5.7. A plot of these errors for each value of d_p can be found in Figures 5.4 through 5.6.

A study of the tables and figures indicates that as d_p increases, the linear models yield optimal discrimination over a wider range of values (for d_r). That is, the critical bounds on d_r such that the use of the first and LDF result in theoretical errors which are essentially the same as the optimum error become wider as the difference in the marginal probabilities increase; this implies that with large mean differences most procedures do well. Table 5.8 displays the critical bounds for each d_p value together with the equation for the optimum error as a function of d_r.

It is also apparent that both the optimum and theoretical errors decrease as d_p increases. As the difference in the marginal probabilities increase, the underlying distributions lie farther apart and therefore discrimination should improve. A somewhat more interesting point which again surfaces is that for all values of d_r

$$\alpha_{-d_r} < \alpha_{d_r}$$

and

$$\alpha[*]_{-d_r} < \alpha[*]_{d_r}$$

where $-d_r$ means $d_r < 0$. Hence discrimination is always better when correlations are negative rather than positive.

For all value of d_r beyond the critical bounds specified (e.g., Table 5.8) the use of either the first or LDF

TABLE 5.5

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.2 WITH p_{1j}=.2,p_{2j}=.4;j=1,2,....6

d_r [r_2(13)]	Optimum Error	Theoretical First	Theoretical LDF
-.6	.2487	.2734	.2616
-.5	.2614	.2760	.2683
-.4	.2710	.2786	.2786
-.3	.2787	.2812	.2812
-.2	.2828	.2838	.2838
-.1	.2864	.2864	.2864
0	.2890	.2890	.2890
.1	.2916	.2916	.2916
.2	.2941	.2941	.2941
.3	.2893	.2967	.3036
.4	.2836	.2993	.3077
.5	.2779	.3019	.3119
.6	.2722	.3045	.3160

TABLE 5.6

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.4
WITH p_{1j}=.3,p_{2j}=.6;j=1,2,....6

d_r [r_2(13)]	Optimum Error	Theoretical First	Theoretical LDF
-.6	.1850	.2036	.1879
-.5	.1942	.2059	.1943
-.4	.2008	.2082	.2082
-.3	.2072	.2105	.2105
-.2	.2128	.2128	.2128
-.1	.2151	.2151	.2151
0	.2174	.2174	.2174
.1	.2197	.2197	.2197
.2	.2221	.2221	.2221
.3	.2244	.2244	.2244
.4	.2209	.2267	.2267
.5	.2149	.2290	.2290
.6	.2089	.2313	.2313

TABLE 5.7

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.4
WITH p_{1j}=.2,p_{2j}=.6;j=1,2,....6

d_r [r_2(13)]	Optimum Error	Theoretical First	Theoretical LDF
-.6	.1198	.1252	.1306
-.5	.1262	.1275	.1326
-.4	.1298	.1298	.1298
-.3	.1321	.1321	.1321
-.2	.1344	.1344	.1344
-.1	.1367	.1367	.1367
0	.1390	.1390	.1390
.1	.1413	.1413	.1413
.2	.1436	.1436	.1436
.3	.1460	.1460	.1460
.4	.1483	.1483	.1483
.5	.1506	.1506	.1506
.6	.1495	.1520	.1529

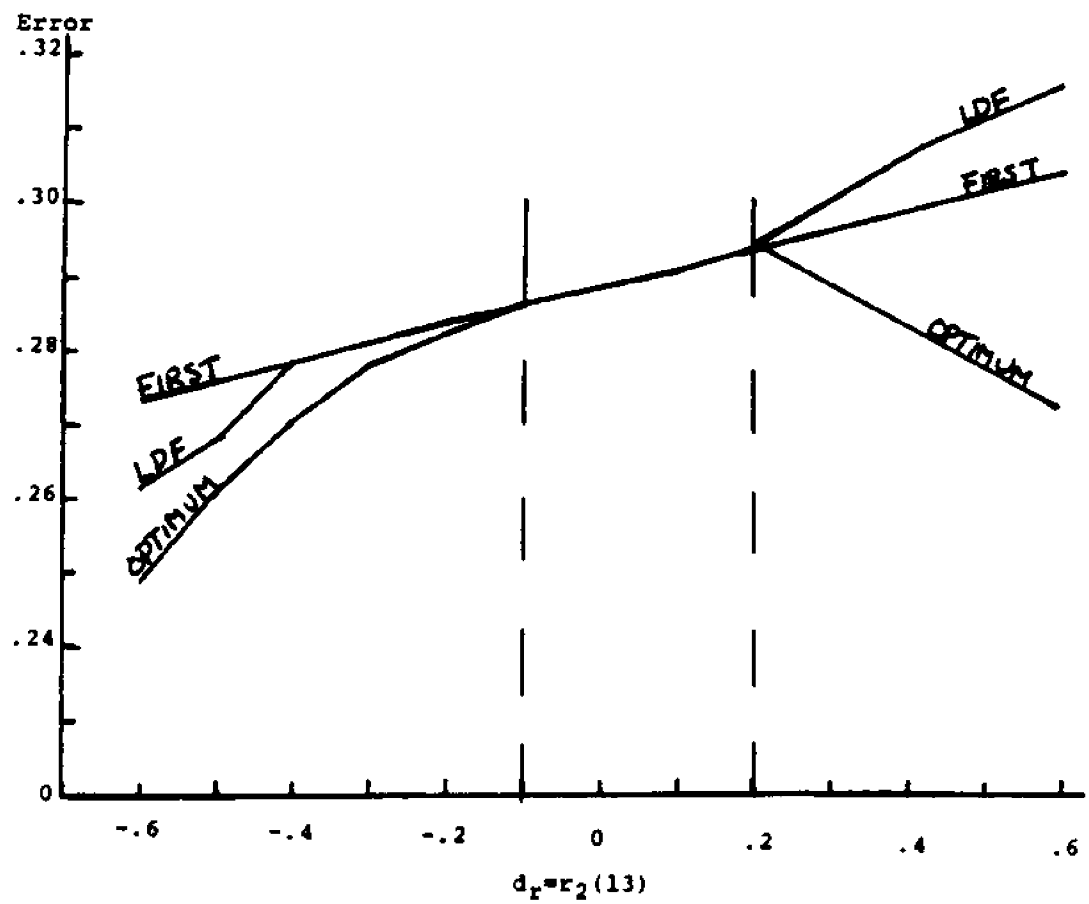

Figure 5.4 - Optimum and Theoretical Errors for $d_p = .2$ with $p_{1j} = .2$, $p_{2j} = .4$; $j = 1, 2, \ldots, 6$, and $d_r = r_2(13)$

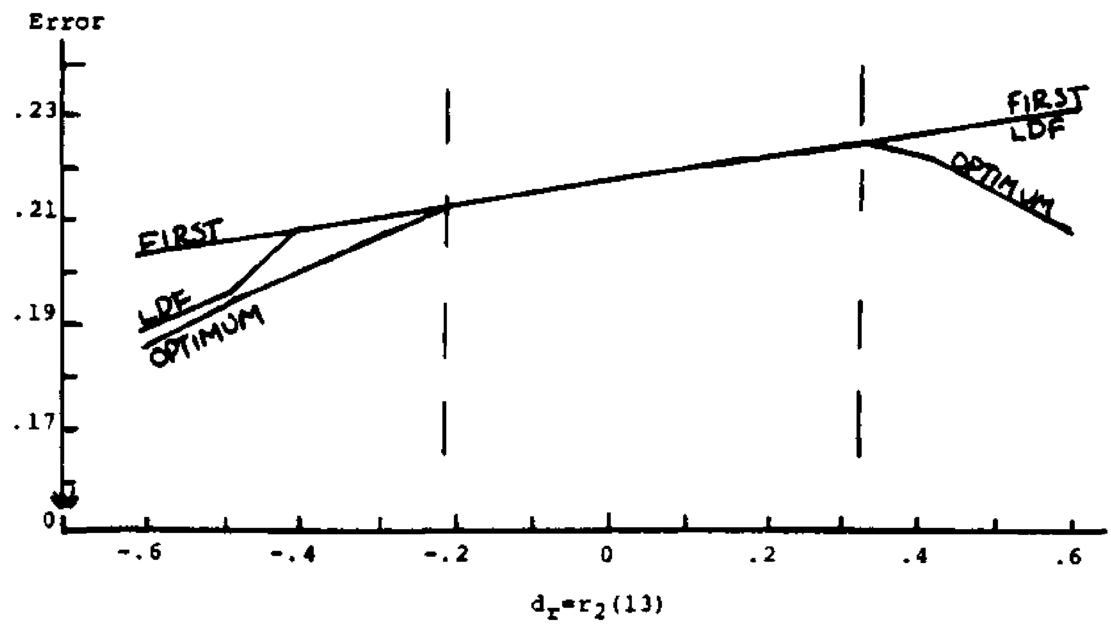

Figure 5.5 - Optimum and Theoretical Errors for $d_p = .3$ with $p_{1j} = .3$, $p_{2j} = .6; j = 1, 2, \ldots\ldots, 6$, and $d_r = r_2(13)$.

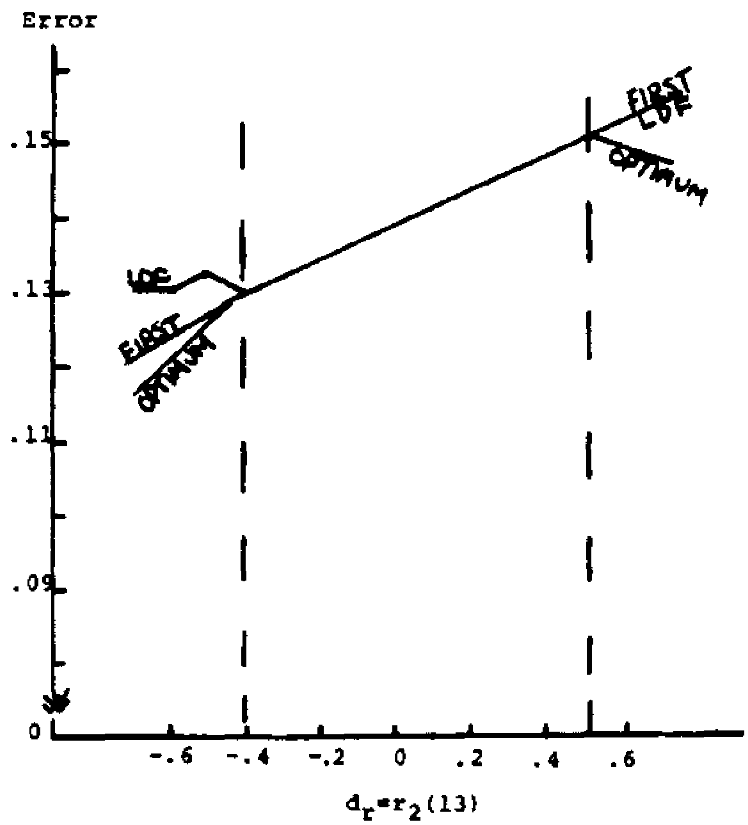

Figure 5.6 - Optimum and Theoretical Errors for $d_p = .4$ with $p_{1j} = .2$, $p_{2j} = .6$; $j = 1, 2, \ldots\ldots, 6$ and $d_r = r_2(13)$.

procedures result in theoretical errors which may be substantially greater than the optimum error. This is especially true for $d_r > d_u$ since over this range the theoretical errors monotonically increase, while the optimum error reverses direction. Differences between the theoretical errors and the optimum error are shown in Table 5.9. It is apparent from the table that as d_p increases, both $[\alpha[1]-\alpha]$, and $[\alpha[\ell]-\alpha]$ decrease. In fact, if $d_p = .4$ then the differences are quite small even for large positive values of d_r. Thus, the evidence suggests that if mean differences are large, then the use of a seemingly inappropriate model does not produce severe anomalies since performance seems relatively unaffected by the correlation structure.

The entries in Tables 5.10 through 5.12 correspond to the mean increases in actual error over optimum error for each value of d_p. A study of the tables indicates that as the sample size is increased, the mean increase in actual error decreases for fixed d_r and d_p. Uniformly, if $d_p \geq .2$ and d_r is fixed, then $\bar{\Lambda}[*]_{n=400} < \bar{\Lambda}[*]_{n=200}$. Recall, analogous results were obtained for population pairs with $d_p = .1$. Also, the tables indicate that the mean increase in actual error is a decreasing function of d_p. Note the linear models seem particularly sensitive to large mean differences. For example, consider population pairs with $d_r = .6$. Here, as d_p went from .2 to .4, there was a 90.1 percent decrease in $\bar{\Lambda}[1]$, while the second largest decrease was 89.6 percent

TABLE 5.8

CRITICAL BOUNDS ON d_r FOR $\alpha[1]= [\ell] =\alpha$
AND d_p=.1,.2,.3, AND .4

d_p	$\alpha[1]=\alpha[\ell]=\alpha$	d_L				d_u
.1	.3997 + .0156 d_r	−.1	$\leq$	d_r	$\leq$	.1
.2	.2890 + .0256 d_r	−.1	$\leq$	d_r	$\leq$	.2
.3	.2174 + .0232 d_r	−.2	$\leq$	d_r	$\leq$	.3
.4	.1390 + .0231 d_r	−.4	$\leq$	d_r	$\leq$	.5

TABLE 5.9

DIFFERENCE BETWEEN THEORETICAL AND OPTIMUM ERROR (IN PERCENT) FOR d_p=.2,.3, AND 4

d_p	d_r [$r_2(13)$]	Optimum Error α	Theoretical First $\alpha[1]-\alpha$	Theoretical LDF $\alpha[\ell]-\alpha$
.2	-.6	.2486	2.47	1.29
	-.5	.2614	1.46	0.69
	-.4	.2710	0.76	0.76
	-.3	.2787	0.25	0.25
	-.2	.2828	0.10	0.10
	-.1	.2864	0	0
	0	.2890	0	0
	.1	.2916	0	0
	.2	.2941	0	0
	.3	.2893	0.74	1.43
	.4	.2836	1.57	2.41
	.5	.2776	2.40	3.40
	.6	.2722	3.23	4.38
.3	-.6	.1850	1.86	0.29
	-.5	.1942	1.17	0.01
	-.4	.2008	0	0
	-.3	.2072	0	0
	-.2	.2128	0	0
	-.1	.2151	0	0
	0	.2174	0	0
	.1	.2197	0	0
	.2	.2221	0	0
	.3	.2244	0	0
	.4	.2209	0.58	0.58
	.5	.2149	1.41	1.41
	.6	.2089	2.24	2.24
.4	-.6	.1198	0.53	1.08
	-.5	.1262	0.13	0.64
	-.4	.1298	0	0
	-.3	.1321	0	0
	-.2	.1344	0	0
	-.1	.1367	0	0
	0	.1390	0	0
	.1	.1413	0	0
	.2	.1436	0	0
	.3	.1460	0	0
	.4	.1483	0	0
	.5	.1506	0	0
	.6	.1495	0.25	0.34

TABLE 5.10

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR $d_p=.2$ WITH $p_{1j}=.2, p_{2j}=.4; j=1,2,\ldots,6$

d_r $[r_2(13)]$	n	Optimum Error	Full	First	Sec-ond	LDF	Matu-sita
	200	.2487	3.42	3.54	1.54	2.42	3.58
-.6	400		1.57	3.05	0.87	1.78	1.61
	∞		0	2.47	0	1.29	0
	200	.2614	3.47	2.42	1.31	1.69	3.86
-.5	400		1.78	1.95	0.70	1.33	1.96
	∞		0	1.46	0	0.69	0
	200	.2710	3.66	1.54	1.50	1.31	4.04
-.4	400		1.85	1.29	0.74	0.92	2.00
	∞		0	0.71	0	0.71	0
	200	.2787	3.79	1.09	1.27	1.00	4.44
-.3	400		1.84	0.71	0.61	0.61	2.15
	∞		0	0.25	0	0.25	0
	200	.2828	4.29	0.84	1.59	1.01	4.73
-.2	400		2.08	0.45	0.75	0.58	2.43
	∞		0	0.10	0	0.10	0
	200	.2864	4.30	0.69	1.62	0.99	4.85
-.1	400		2.20	0.40	0.79	0.61	2.56
	∞		0	0	0	0	0
	200	.2890	4.50	0.66	1.66	0.88	5.00
0	400		2.24	0.31	0.85	0.54	2.57
	∞		0	0	0	0	0
	200	.2910	4.41	0.57	1.60	0.84	4.83
.1	400		2.05	0.31	0.79	0.54	2.36
	∞		0	0	0	0	0
	200	.2941	4.01	0.54	1.26	0.71	4.43
.2	400		1.73	0.29	0.51	0.43	2.04
	∞		0	0	0	0	0
	200	.2893	3.97	1.18	1.65	1.24	4.38
.3	400		2.12	0.98	0.81	1.09	2.38
	∞		0	0.74	0	1.43	0
	200	.2836	4.38	2.09	2.02	2.05	4.73
.4	400		2.25	1.81	0.95	1.93	2.46
	∞		0	1.57	0	2.41	0
	200	.2779	4.27	2.80	2.10	2.76	4.64
.5	400		2.30	2.64	1.01	2.52	2.64
	∞		0	2.40	0	3.40	0
	200	.2722	3.69	3.67	1.88	3.23	3.90
.6	400		1.83	3.47	0.94	3.18	1.99
	∞		0	3.23	0	4.38	0

TABLE 5.11

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR $d_p=.3$ WITH $p_{1j}=.3, p_{2j}=.6; j=1,2,\ldots,6$

d_r [r_2(13)]	n	Optimum Error	Full	First	Sec-ond	LDF	Matu-sita
	200	.1850	3.01	2.49	1.06	1.86	3.01
-.6	400		1.51	2.06	0.47	1.13	1.53
	∞		0	1.86	0	0.29	0
	200	.1942	3.11	1.66	1.06	1.27	3.10
-.5	400		1.72	1.28	0.45	0.77	1.72
	∞		1.17	0	0	0.01	0
	200	.2008	3.42	1.20	1.30	1.18	3.42
-.4	400		1.74	0.84	0.55	0.71	1.74
	∞		0	0	0	0	0
	200	.2072	3.39	0.84	1.31	1.03	3.39
-.3	400		1.76	0.44	0.55	0.47	1.75
	∞		0	0	0	0	0
	200	.2182	3.38	0.36	1.16	0.59	3.39
-.2	400		1.89	0.11	0.45	0.20	1.92
	∞		0	0	0	0	0
	200	.2151	3.96	0.34	1.51	0.63	3.97
-.1	400		2.12	0.11	0.73	0.27	2.15
	∞		0	0	0	0	0
	200	.2174	3.81	0.32	1.58	0.64	3.82
0	400		2.12	0.08	0.77	0.29	2.12
	∞		0	0	0	0	0
	200	.2197	3.90	0.36	1.63	0.67	3.91
.1	400		2.05	0.11	0.74	0.27	2.08
	∞		0	0	0	0	0
	200	.2221	3.38	0.33	1.37	0.61	3.36
.2	400		1.78	0.07	0.65	0.27	1.81
	∞		0	0	0	0	0
	200	.2244	2.95	0.35	1.07	0.64	2.97
.3	400		1.53	0.07	0.42	0.31	1.53
	∞		0	0	0	0	0
	200	.2209	3.15	0.81	1.26	1.21	3.17
.4	400		1.70	0.61	0.57	1.02	1.71
	∞		0	0.58	0	0.58	0
	200	.2149	3.34	1.74	1.43	2.04	3.34
.5	400		1.93	1.51	0.78	1.77	1.94
	∞		0	1.41	0	1.41	0
	200	.2089	3.39	2.58	1.50	2.82	3.31
.6	400		1.92	2.34	0.94	2.85	1.90
	∞		0	2.24	0	2.24	0

TABLE 5.12

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR d_P=.4 WITH p_{1j}=.2, p_{2j}=.6, j=1,2,....,6

d_r [r_2(13)]	n	Optimum Error	Full	First	Second	LDF	Matusita
	200	.1198	1.93	0.68	0.68	0.95	1.94
-.6	400		0.84	0.56	0.48	0.98	0.83
	∞		0	0.54	0	1.08	0
	200	.1262	2.08	0.25	0.77	0.59	2.10
-.5	400		0.94	0.15	0.55	0.46	0.95
	∞		0	0.13	0	0.64	0
	200	.1298	2.43	0.11	0.95	0.36	2.46
-.4	400		1.05	0.01	0.54	0.24	1.06
	∞		0	0	0	0	0
	200	.1321	2.75	0.11	1.07	0.35	2.76
-.3	400		1.40	0.03	0.50	0.12	1.39
	∞		0	0	0	0	0
	200	.1344	3.05	0.13	1.14	0.20	3.07
-.2	400		1.55	0.02	0.54	0.05	1.56
	∞		0	0	0	0	0
	200	.1367	2.95	0.17	1.15	0.22	2.95
-1	400		1.50	0.01	0.47	0.02	1.49
	∞		0	0	0	0	0
	200	.1390	3.30	0.14	1.14	0.30	3.30
0	400		1.59	0.01	0.49	0.03	1.57
	∞		0	0	0	0	0
	200	.1413	3.17	0.18	1.19	0.33	3.18
.1	400		1.55	0.01	0.55	0.04	1.53
	∞		0	0	0	0	0
	200	.1436	3.08	0.17	1.29	0.32	3.10
.2	400		1.69	0.03	0.57	0.06	1.72
	∞		0	0	0	0	0
	200	.1460	2.90	0.11	1.21	0.30	2.91
.3	400		1.50	0.01	0.56	0.06	1.50
	∞		0	0	0	0	0
	200	.1483	2.61	0.10	0.99	0.30	2.64
.4	400		1.29	0	0.49	0.10	1.29
	∞		0	0	0	0	0
	200	.1506	2.05	0.10	0.80	0.28	2.11
.5	400		1.06	0	0.46	0.05	1.12
	∞		0	0	0	0	0
	200	.1495	1.96	0.46	0.91	0.77	2.00
.6	400		0.91	0.34	0.51	0.33	0.96
	∞		0	0.76	0	0.76	0

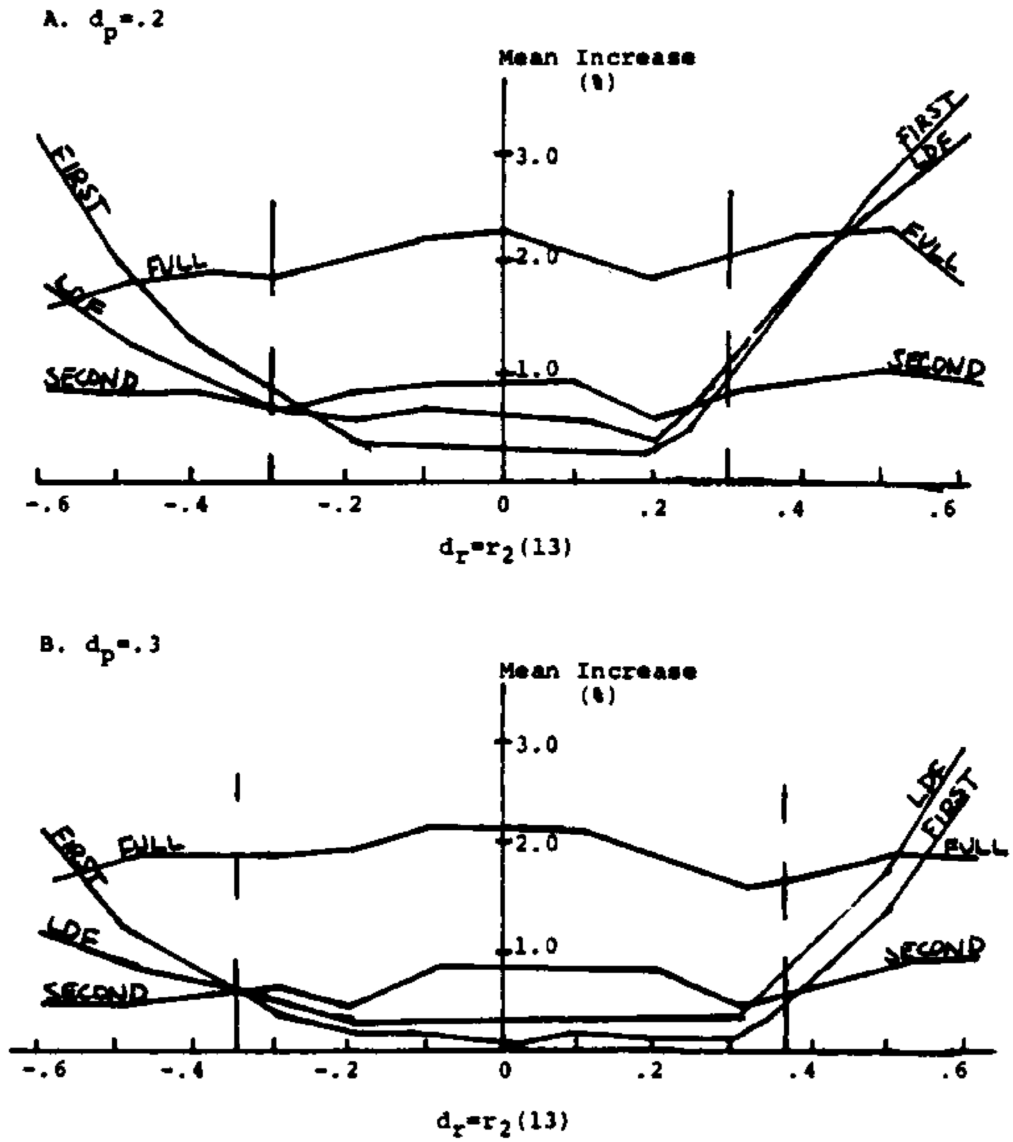

Figure 5.7 - Mean Increase in Actual Over Optimum Error (in Percent) Based on 100 Monte Carlo Trials for d_p=.2,.3, and .4, n=400 and $d_r=r_2(13)$.

C. $d_p = .4$

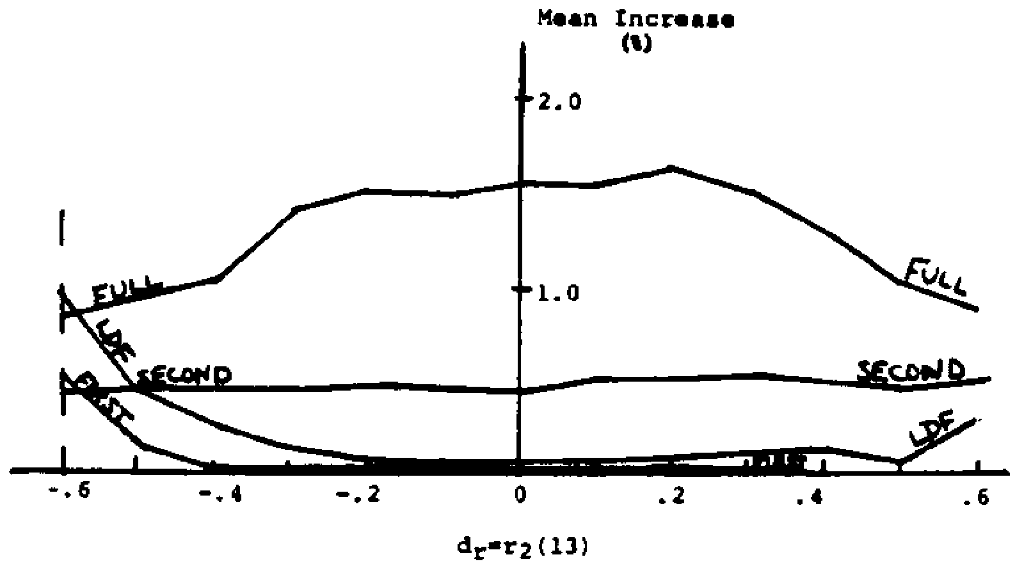

Figure 5.7 (continued)

in $\bar{\Lambda}[\ell]$. The respective decreases for $\bar{\Lambda}[f]$, $\bar{\Lambda}[2]$ and $\bar{\Lambda}[M]$ were 50.2 percent, 45.7 percent, and 51.7 percent. Figure 5.7, which plots the mean increases for each discrimination procedure at n=400, clearly supports this contention.

The vertical dashed-lines in each figure indicate the critical lower and upper bounds on d_r such that for values of d_r within these limits, the use of the linear models yield lower mean increases in actual error than any of the other procedures. The sensitivity of the linear models to large mean differences is again demonstrated by the widening of these bounds as d_p increases. In fact, when d_p=.4, the linear models "flatten out" such that their performance is superior to the other procedures over most values of d_r; i.e., note the absence of any critical upper bound on d_r when d_p=.4.

The mean correlations between the true L.L.R.'s and the observed L.L.R.'s are displayed in Tables 5.13 through 5.15. For fixed d_p the mean correlations again exhibit a good deal of non-monotonicity as d_r is varied. However, a study of the tables yields the following general remarks:

1. Even when the tone state probabilities are known, neither the first nor the LDF can lead to optimal discrimination whenever $d_r \neq 0$.

TABLE 5.13

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR d_p=.2 WITH p_{1j}=.2, p_{2j}=.4; j=1,2....6

d_r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
[r_2(13)]							
	200	.7174	.6243	.7491	.5810	.7610	.130
-.6	400	.6609	.6196	.7326	.5288	.7264	.083
	∞	1.0	.6149	1.0	.5172	1.0	
	200	.5852	.8182	.6162	.7720	.6384	.091
-.5	400	.4272	.8087	.7044	.7510	.5274	.037
	∞	1.0	.8158	1.0	.7632	1.0	
	200	.4454	.8671	.5314	.8175	.4793	.060
-.4	400	.6234	.9001	.8445	.8904	.6773	.021
	∞	1.0	.9051	1.0	.8774	1.0	
	200	.4961	.9273	.8175	.9365	.4829	.053
-.3	400	.6264	.9545	.8593	.9450	.6275	.015
	∞	1.0	.9540	1.0	.9408	1.0	
	200	.4767	.9813	.9131	.9755	.5096	.046
-.2	400	.6497	.9661	.8840	.9512	.6370	.012
	∞	1.0	.9817	1.0	.9766	1.0	
	200	.5652	.9756	.7157	.9798	.5520	.048
-.1	400	.7614	.9827	.8687	.9869	.6932	.014
	∞	1.0	.9958	1.0	.9947	1.0	
	200	.5117	.9845	.7872	.9674	.4711	.072
0	400	.6232	.9792	.8316	.9669	.6129	.018
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.5734	.9762	.7183	.9712	.5385	.054
.1	400	.7539	.9741	.8830	.9737	.7212	.014
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.5782	.9799	.9359	.9736	.5874	.041
.2	400	.6649	.9616	.8561	.9618	.6310	.012
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.5688	.9189	.8557	.8451	.5390	.054
.3	400	.5756	.9613	.8716	.9495	.5541	.013
	∞	1.0	.9680	1.0	.9620	1.0	
	200	.4222	.9244	.5115	.9357	.4409	.040
.4	400	.7153	.9377	.9308	.9052	.7150	.012
	∞	1.0	.9438	1.0	.9344	1.0	
	200	.6433	.9031	.8006	.8916	.6706	.065
.5	400	.5751	.8992	.9118	.8788	.6179	.018
	∞	1.0	.9124	1.0	.8994	1.0	
	200	.5543	.8541	.8398	.8264	.6008	.067
.6	400	.6273	.8633	.9222	.8462	.6755	.018
	∞	1.0	.8723	1.0	.8560	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

TABLE 5.14

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR d_p=.3 WITH p_{1j}=.3, p_{2j}=.6; j=1,2....6

d_r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
$[r_2(13)]$							
	200	.8122	.8412	.9020	.8729	.8643	.100
-.6	400	.8909	.8641	.9375	.8913	.9035	.074
	∞	1.0	.8711	1.0	.9001	1.0	
	200	.7641	.9132	.8880	.9308	.8200	.058
-.5	400	.8312	.9186	.9278	.9333	.8719	.024
	∞	1.0	.9255	1.0	.9377	1.0	
	200	.8318	.9396	.9443	.9547	.8616	.030
-.4	400	.8378	.9552	.9493	.9610	.8703	.007
	∞	1.0	.9564	1.0	.9617	1.0	
	200	.7793	.9627	.8759	.7681	.8440	.015
-.3	400	.7372	.9680	.9687	.9707	.7853	.002
	∞	1.0	.9768	1.0	.9788	1.0	
	200	.6356	.9794	.8358	.9722	.7517	.013
-.2	400	.7997	.9814	.9573	.9777	.8283	.001
	∞	1.0	.9900	1.0	.9907	1.0	
	200	.7329	.9780	.9511	.9710	.8115	.008
-.1	400	.8483	.9873	.9703	.9892	.8542	-
	∞	1.0	.9976	1.0	.9977	1.0	
	200	.6899	.9892	.8235	.9788	.7587	.008
0	400	.7927	.9898	.9634	.9898	.8273	-
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.7468	.9822	.9490	.9798	.8172	.007
.1	400	.8311	.9874	.9612	.9887	.8673	-
	∞	1.0	.9975	1.0	.9975	1.0	
	200	.6254	.9760	.8436	.9783	.7453	.011
.2	400	.8072	.9812	.9766	.9834	.8463	.001
	∞	1.0	.9902	1.0	.9904	1.0	
	200	.7041	.9532	.9239	.9559	.7845	.009
.3	400	.7659	.9792	.9683	.9633	.8227	.001
	∞	1.0	.9777	1.0	.9778	1.0	
	200	.7512	.9545	.9367	.9487	.8273	.012
.4	400	.7785	.9517	.9764	.9459	.8438	.001
	∞	1.0	.9594	1.0	.9595	1.0	
	200	.6361	.9126	.8599	.9134	.7581	.022
.5	400	.8001	.9217	.9707	.9211	.8514	.007
	∞	1.0	.9344	1.0	.9344	1.0	
	200	.8161	.8710	.8887	.8700	.8894	.037
.6	400	.7484	.8730	.8744	.8720	.8163	.010
	∞	1.0	.9011	1.0	.9004	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

TABLE 5.15

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR d_p=.1 WITH p_{1j}=.4, p_{2j}=.5; j=1,2....6

d_r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
$[r_2(13)]$							
	200	.9139	.9163	.9093	.9360	.9211	.130
-.6	400	.8720	.9112	.9286	.9341	.8913	.081
	∞	1.0	.9255	1.0	.9409	1.0	
	200	.8277	.9541	.8676	.9559	.8703	.096
-.5	400	.8123	.9564	.9174	.9557	.8598	.036
	∞	1.0	.9770	1.0	.9639	1.0	
	200	.8426	.9765	.8762	.9775	.8616	.059
-.4	400	.8074	.9746	.9832	.9767	.8665	.019
	∞	1.0	.9770	1.0	.9781	1.0	
	200	.7968	.9757	.8698	.9789	.8642	.055
-.3	400	.8347	.9848	.9736	.9884	.8527	.013
	∞	1.0	.9881	1.0	.9881	1.0	
	200	.8102	.9901	.9765	.9899	.8563	.046
-.2	400	.8594	.9938	.9694	.9927	.8439	.013
	∞	1.0	.9950	1.0	.9948	1.0	
	200	.7482	.9891	.9087	.9845	.7860	.048
-.1	400	.8634	.9935	.9694	.9956	.9047	.014
	∞	1.0	.9988	1.0	.9987	1.0	
	200	.7320	.9925	.8735	.9743	.8155	.053
0	400	.8005	.9986	.9358	.9967	.7990	.018
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.7384	.9777	.9187	.9651	.7675	.048
.1	400	.8307	.9943	.9720	.9909	.8978	.014
	∞	1.0	.9988	1.0	.9987	1.0	
	200	.7784	.9936	.9623	.9714	.8672	.044
.2	400	.8122	.9924	.9696	.9906	.8685	.012
	∞	1.0	.9952	1.0	.9948	1.0	
	200	.7714	.9826	.9329	.9685	.8417	.047
.3	400	.8150	.9880	.9672	.9841	.8538	.013
	∞	1.0	.9891	1.0	.9881	1.0	
	200	.7749	.9747	.8449	.9754	.8442	.045
.4	400	.8152	.9755	.9667	.9617	.8773	.013
	∞	1.0	.9799	1.0	.9782	1.0	
	200	.7581	.9638	.8886	.9540	.8293	.058
.5	400	.7836	.9649	.9471	.9561	.7963	.017
	∞	1.0	.9670	1.0	.9645	1.0	
	200	.7881	.9424	.8646	.9197	.8527	.076
.6	400	.7634	.9467	.9376	.9382	.8553	.019
	∞	1.0	.9490	1.0	.9456	1.0	

[a]The average proportion of times π_i=(x;[2]) was less than zero.

2. The behavior of the mean correlations were somewhat erratic with respect to increased sample sizes. For all values of d_p and fixed d_r, it was expected that $\rho[*]_{n=400} > \rho[*]_{n=200}$; however, this was not universally true.

3. For moderate values of d_r, the performance of the linear models was always better than that of the full, second or Matusita. With large mean differences, the superiority of the first and LDF procedures extended over a wider range of values (for d_r).

4. In most cases, for fixed d_r, an increase in d_p resulted in higher mean correlations for each discrimination procedure.

5.11.1 - Conclusions: Sampling Under Case (i): The population pairs considered under Case (i) were of a rather simplistic nature in that they contained only one non-zero correlation term. Nevertheless, the results did indicate that performance (of each discrimination procedure) was affected as the non-zero correlation term was varied.

What follows is a summary of the major findings derived from the initial series of sampling experiments:

1. It was always possible to determine critical bounds on d_r, such that if they are exceeded, the use of the first or LDF procedures will result in greater misclassification error (on the average) than either the full, second or

Matusita. For small mean differences, these critical bounds are likely to be more restricted.

2. The evidence suggests that highly correlated variables may yield better discrimination than uncorrelated variables. Also, for fixed d_p, negative correlations were found to improve discrimination across all procedures. That is, with negative correlations both the optimum error and the theoretical errors were lower than with positive correlations.

3. The linear discrimination procedures seem to be most sensitive to large mean differences. For large d_p, the overall performance of both the first and LDF procedures substantially improved and their performances were not significantly different from the other procedures. Thus, with large d_p, all of these procedures do well and hence the choice of a particular discrimination procedure is no longer a serious problem.

<u>5.12 - Case (ii)</u>: The population pairs considered under this case are of a more complex nature. Here, all off-diagonal correlations can take on non-zero values, but it is assumed that $r_1(jk) \neq r_2(jk)$, for all $j \neq k$, $j,k=1,2....,6$. The correlation structures initially examined are described in Tables 5.16. For these population pairs $d_r[r_2(jk)-r_1(jk)]$ ranged from -.15 to .23. At first the difference between the marginal probability vectors in the two populations, again denoted by d_p, was taken to be .1 with $p_{1j}=.4$

TABLE 5.16

CASE (ii) CORRELATION STRUCTURES

Population 1 | Population 2

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & .10 & \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix} \qquad \begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & -.05 & \\ & & 1.0 & & & \\ -.05 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & .10 & \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix} \qquad \begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & 0 & \\ & & 1.0 & & & \\ 0 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & .10 & \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix} \qquad \begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & .05 & \\ & & 1.0 & & & \\ & & & 1.0 & & \\ .05 & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & .10 & \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix} \qquad \begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & .10 & \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

Table 5.16 (continued)

Population 1

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & & .10 \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & & .10 \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & & .10 \\ & & 1.0 & & & \\ .10 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

Population 2

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & & .20 \\ & & 1.0 & & & \\ .20 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & & .30 \\ & & 1.0 & & & \\ .30 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & & & & & \\ & 1.0 & & & & .33 \\ & & 1.0 & & & \\ .33 & & & 1.0 & & \\ & & & & 1.0 & \\ & & & & & 1.0 \end{bmatrix}$$

and $p_{2j}=.5$ for $j=1,2,....,6$. Summary results for the optimum error and the theoretical errors are shown in Table 5.17 and Figure 5.8.

From the table and figure it is apparent that the critical bounds on d_r such that the linear procedures yield optimal error are quite narrow, i.e., $-.05\leq d_r\geq 0$. For values of d_r beyond the critical limits specified, the use of either the first or LDF procedures may result in theoretical errors which are significantly greater than the optimum error. Significantly greater misclassification error occurs with the use of the linear models if $d_r>0$. On the other hand, the increase in theoretical error over optimum error is far less severe if $d_r<-.05$ since over this range both theoretical and optimum errors are monotonically decreasing functions.

Table 5.18 and Figure 5.9 present summary results for the mean increase in actual over optimum error for each discrimination procedure. In prior experiments the effect of larger sample sizes has been to reduce the mean increase in actual error across all procedures. Although this is again true for the full, second and Matusita procedures, a study of the corresponding behavior of the linear models yields a different conclusion. If $d_r>0$, then both $\bar{\pi}[1]$ and $\bar{\pi}[\ell]$ increase as n goes from 200 to 400. Note the increases are for the most part slight. However, this does not alter the conclusion that there may be cases where discrimination for some procedure is improved by limiting the sample size rather than by increasing it.

TABLE 5.17

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.1
WITH p_{1j}=.4,p_{2j}=.5;j=1,2,....6

d_r $[r_2(jk)-r_1(jk)]$	Optimum Error	Theoretical First	Theoretical LDF
-.15	.3752	.3880	.3880
-.10	.3992	.3997	.3997
-.05	.4114	.4114	.4114
0	.4232	.4232	.4232
.10	.4054	.4466	.4466
.20	.3736	.4700	.4700
.23	.3575	.4771	.4771

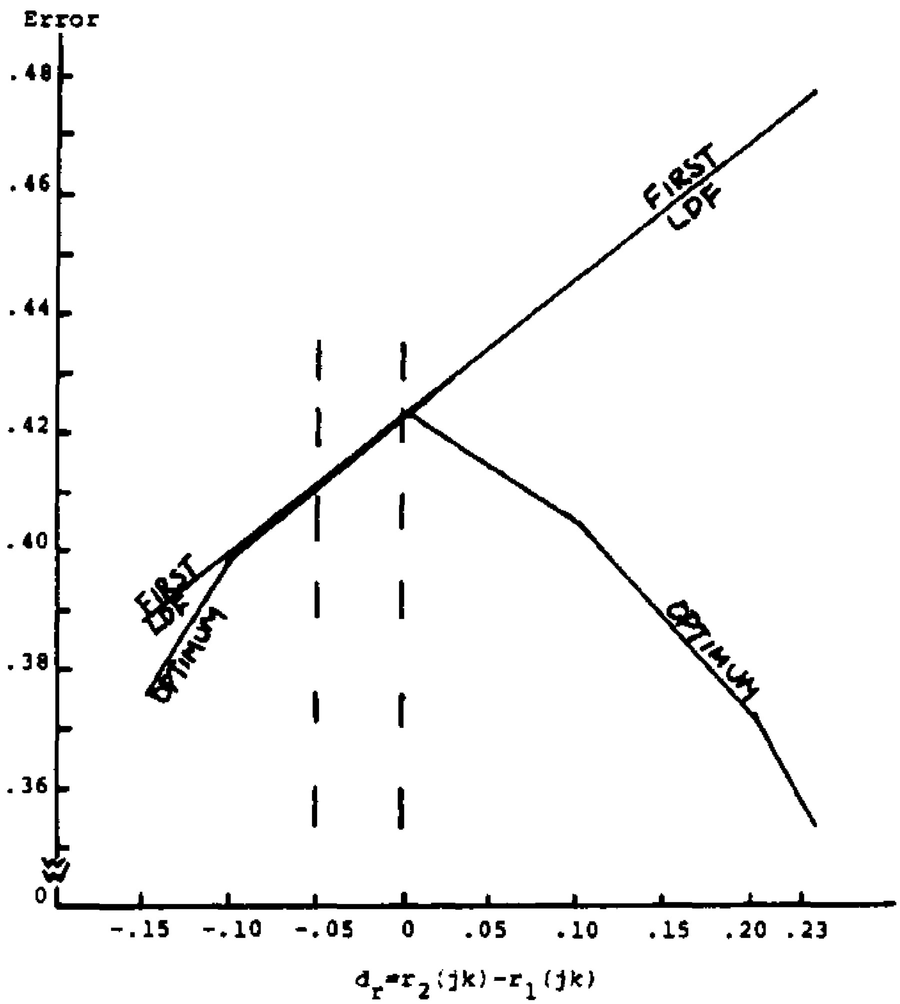

Figure 5.8 - Optimum and Theoretical Errors for d_p=.1 with p_{1j}=.4, p_{2j}=.5; j=1,2,......,6 and $r_1(jk) \neq r_2(jk)$.

TABLE 5.18

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT)
BASED ON 100 MONTE CARLO TRIALS FOR
d_p=.1 WITH p_{1j}=.4,p_{2j}=.5,j=1,2,....6

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Full	First	Sec-ond	LDF	Matu-sita
	200	.3752	4.17	3.52	2.12	3.87	4.17
-.5	400		2.87	3.08	1.20	3.32	2.89
	∞		0	1.28	0	1.28	0
	200	.3992	4.05	1.73	2.51	2.19	4.05
-.10	400		2.95	1.41	1.44	1.79	2.96
	∞		0	0.05	0	0.05	0
	200	.4114	4.04	1.05	2.64	1.53	4.04
-.05	400		3.03	0.65	1.67	1.02	3.05
	∞		0	0	0	0	0
	200	.4232	3.46	0.42	2.21	0.92	3.47
0	400		2.63	0.18	1.46	0.44	2.67
	∞		0	0	0	0	0
	200	.4054	4.28	3.20	2.96	3.41	4.25
.10	400		3.15	3.27	2.22	3.45	3.06
	∞		0	4.12	0	4.12	0
	200	.3736	3.22	7.47	1.91	7.20	3.11
.20	400		2.10	7.79	1.18	7.76	2.05
	∞		0	9.64	0	9.64	0
	200	.3575	2.98	9.45	1.88	8.99	2.87
.23	400		1.79	9.90	1.18	9.40	1.72
	∞		0	11.96	0	11.96	0

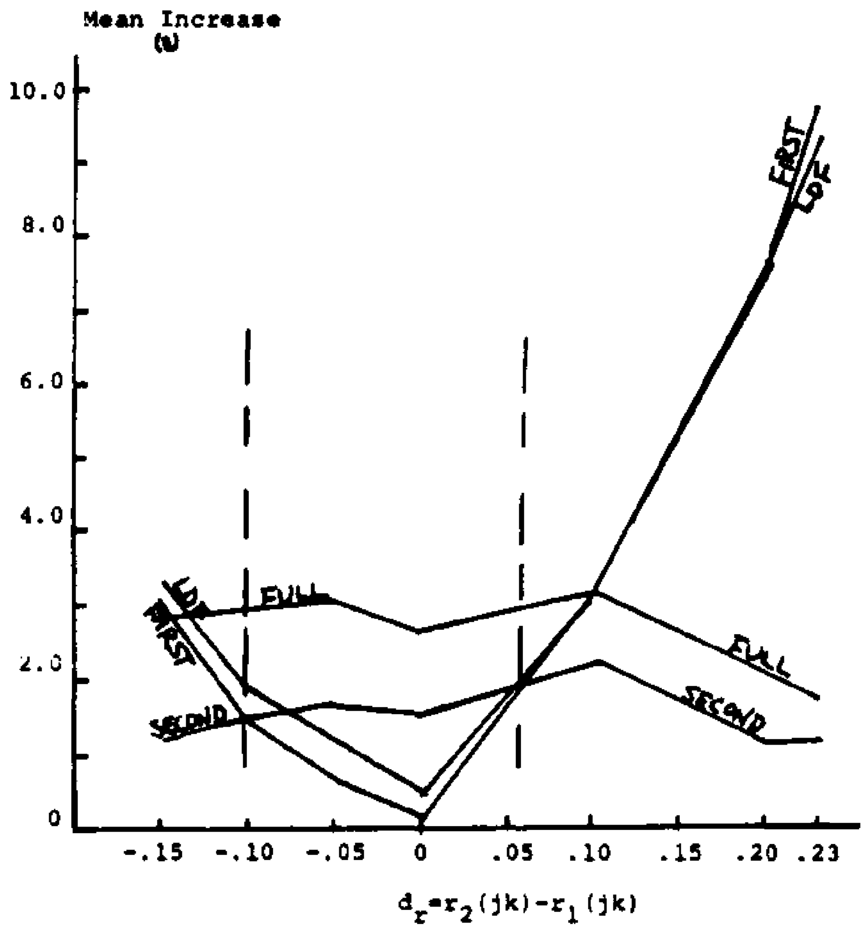

Figure 5.9 - Mean Increase in Actual Over Optimum Error (in Percent) Based on 100 Monte Carlo Trials for d_p=.1 with p_{1j}=.4, p_{2j}=.5; j=1,2,.....,6, n=400 and $r_1(jk) \neq r_2(jk)$.

It is also apparent that the performance of the linear models is quite good if $-.10 \leq d_r \leq .05$. Within this range, the use of the linear models results in smaller mean increases in actual error than any other procedure. On the other hand, as d_r becomes large the performance of the first and LDF procedures falls off quite rapidly.

The mean correlations for each discrimination procedure are shown in Table 5.19. For $d_r \geq .20$, the use of the full, second or Matusita procedures results in significantly higher mean correlations than either the first or LDF. The susceptibility of the linear models to population structures containing large correlations is demonstrated by the low mean correlations found in the third row (corresponding to $n= \infty$) for each value of d_r. In other words, even when the true state probabilities are known, the observed L.L.R's for the first and LDF procedures are linearly different from the true L.L.R.'s whenever d_r is large. In addition, a comparison of the mean correlations again show the Matusita procedure to be slightly better than full, but not as good as the second for most values of d_r and n.

The next series of sampling experiments were designed to investigate the effects of increasing the magnitude of the values assigned to all $r_i(jk)$ on the performance of each discrimination procedure. Recall, the definition of d_r is somewhat arbitrary since a specified value for d_r does not uniquely determine the values of either $r_1(jk)$ or $r_2(jk)$.

TABLE 5.19

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR d_p=.1 WITH p_{1j}=.4; p_{2j}=.5, j=1,2,....6

d_r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
$[r_2(jk)-r_1(jk)]$							
	200	.4650	.4470	.7034	.4320	.5944	.006
-.15	400	.5027	.4564	.8304	.4421	.5935	.002
	∞	1.0	.4759	1.0	.4759	1.0	
	200	.3267	.7059	.7097	.6977	.4536	.001
-.10	400	.5694	.6599	.7749	.6199	.5866	-
	∞	1.0	.7145	1.0	.7145	1.0	
	200	.1744	.8706	.5931	.8577	.3145	.001
-.05	400	.4078	.8724	.8034	.8425	.4862	-
	∞	1.0	.9070	1.0	.9070	1.0	
	200	.1677	.8625	.5666	.6939	.2897	.002
0	400	.2847	.9661	.7187	.9441	.3574	-
	∞	1.0	.9789	1.0	.9789	1.0	
	200	.0557	.3905	.2117	.2125	.1154	.015
.10	400	.4192	.6652	.6125	.6375	.4903	.002
	∞	1.0	.6731	1.0	.6731	1.0	
	200	.6892	.2901	.7795	.2346	.7107	.095
.20	400	.7298	.2901	.8031	.2316	.7696	.075
	∞	1.0	.3736	1.0	.3736	1.0	
	200	.7469	.1415	.8067	.0996	.6737	.150
.23	400	.9242	.1423	.7255	.1057	.8929	.148
	∞	1.0	.1524	1.0	.1524	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

Although different values of $r_1(jk)$ and $r_2(jk)$ can be chosen yielding identical d_r values, the effect of varying the magnitude of $r_i(jk)$ on performance is not clear. In order to determine whether results are generalizable across experiments for population pairs containing similar d_r values, two additional sampling experiments were accomplished. In each experiment the values assigned to the $r_1(jk)$ elements were incremented in steps of .10, while the $r_2(jk)$ elements assumed values as shown in Table 5.16. Thus, in the first experiment all $r_2(jk)=.20$, while in the second experiment all $r_2(jk)=.30$. In this way it was possible to talk to population pairs having identical d_r values but different correlation matrices. Tables 5.20 and 5.21 present summary results.

Two main points are revealed from Table 5.20. First, the functional relationship between the theoretical errors and d_r indicates that the rate of change in $\alpha[1]$ and $\alpha[\ell]$ per unit increase in d_r is constant across all population structures. Note for all values of d_r, $\alpha[1]=\alpha[\ell]$ and uniformly $\Delta\alpha[1,\ell]/\Delta d_r=0.245$. Thus, increasing the magnitude of all $r_1(jk)$ correlation terms results in a parallel upward shift in the linear functions. Secondly, for fixed d_r, the behavior of the optimum error is far less exact. Here, the optimum error exhibits a good deal of non-monotonicity as the magnitude of the $r_1(jk)$ correlation terms is varied. However, it is apparent that the difference between the theoretical error (linear procedures) and the optimum error

TABLE 5.20

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.1
WITH p_{1j}=.4,p_{2j}=.5;j=1,2,....6,
AND $r_1(jk)$=.10, .20 AND .30

d_r $[r_2(jk)-r_1(jk)]$	Optimum Error	Theoretical First	Theoretical LDF
All $r_1(jk)$=.10			
-.15	.3752	.3880	.3880
-.10	.3992	.3997	.3997
-.05	.4114	.4114	.4114
0	.4232	.4232	.4232
.10	.4054	.4466	.4466
.20	.3736	.4700	.4700
.23	.3575	.4771	.4771
$\alpha[1]=\alpha[\ell]=.4232+.2345\ d_r$			
All $r_1(jk)$=.20			
-.25	.3133	.3880	.3880
-.20	.3484	.3997	.3997
-.15	.3779	.4114	.4114
-.10	.3954	.4232	.4232
0	.4240	.4466	.4466
.10	.4021	.4700	.4700
.13	.3951	.4771	.4771
$\alpha[1]=\alpha[\ell]=.4466+.2345\ d_r$			
All $r_2(jk)$=.30			
-.35	.2513	.3880	.3880
-.30	.2865	.3997	.3997
-.25	.3216	.4114	.4114
-.20	.3531	.4232	.4232
-.10	.3818	.4466	.4466
0	.3863	.4700	.4700
.03	.3492	.4771	.4771
$\alpha[1]=\alpha[\ell]=.4700+.2345\ d_r$			

TABLE 5.21

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR d_p=.1 WITH p_{1j}=.4,p_{2j}=.5,j=1,2,....,6; AND r_1(jk)=.10,.20, AND .30

d_r [r_2(jk)-r_1(jk)]	n	Optimum Error	Full	First	Second	LDF	Matusita
All r_1(jk)=.10							
	200	.3752	4.17	3.52	2.12	3.87	4.17
-.15	400		2.87	3.08	1.20	3.32	2.89
	∞		0	1.28	0	1.28	0
	200	.3992	4.05	1.73	2.51	2.19	4.05
-.10	400		2.95	1.41	1.44	1.79	2.96
	∞		0	0.05	0	0.05	0
	200	.4114	4.04	1.05	2.64	1.53	4.04
-.05	400		3.03	0.65	1.67	1.02	3.05
	∞		0	0	0	0	0
	200	.4232	3.46	0.42	2.21	0.92	3.47
0	400		2.63	0.18	1.46	0.44	2.67
	∞		0	0	0	0	0
	200	.4054	4.28	3.20	2.96	3.41	4.25
.10	400		3.15	3.27	2.22	3.45	3.06
	∞		0	4.12	0	4.12	0
	200	.3736	3.22	7.47	1.91	7.20	3.11
.20	400		2.10	7.79	1.18	7.76	2.05
	∞		0	9.64	0	9.64	0
	200	.3575	2.98	9.45	1.88	8.99	2.87
.23	400		1.79	9.90	1.18	9.40	1.72
	∞		0	11.96	0	11.96	0
All r_1(jk)=.20							
	200	.3133	5.00	10.37	2.65	10.90	5.03
-.25	400		3.36	9.62	1.29	10.55	3.38
	∞		0	7.47	0	7.47	0
	200	.3484	5.25	7.48	3.00	7.93	5.27
-.20	400		3.29	7.15	1.54	7.64	3.31
	∞		0	5.13	0	5.13	0
	200	.3779	4.52	5.00	2.87	5.46	4.61
-.15	400		3.30	4.66	1.67	5.10	3.48
	∞		0	3.35	0	3.35	0
	200	.3954	4.78	3.86	3.15	4.11	4.78
-.10	400		3.51	3.59	2.06	3.85	3.52
	∞		0	2.78	0	2.78	0
	200	.4240	2.99	1.91	1.86	1.88	2.99
0	400		2.11	2.06	1.31	1.74	2.12
	∞		0	2.26	0	2.26	0
	200	.4021	2.83	5.24	1.83	4.57	2.67
.10	400		1.83	5.64	1.22	4.59	1.74
	∞		0	6.79	0	6.79	0
	200	.3951	2.43	6.22	1.35	5.36	2.27
.13	400		1.11	6.74	0.63	5.41	1.04
	∞		0	8.20	0	8.20	0

TABLE 5.21 (CONTINUED)

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Full	First	Second	LDF	Matusita
All $r_1(jk)=.30$							
	200	.2513	3.78	17.35	1.37	17.72	3.82
-.35	400		1.87	16.63	0.64	17.60	1.95
	∞		0	13.67	0	0	0
	200	.2865	4.32	14.28	1.78	14.50	4.41
-.30	400		2.09	13.94	0.73	14.47	2.16
	∞		0	11.32	0	11.32	0
	200	.3216	4.21	11.20	1.97	11.60	4.31
-.25	400		2.25	10.85	0.86	11.37	2.35
	∞		0	8.98	0	8.98	0
	200	.3531	3.78	8.66	1.77	8.82	4.05
-.20	400		2.03	8.41	0.71	8.63	2.16
	∞		0	7.01	0	7.01	0
	200	.3818	4.65	6.71	3.12	6.36	4.80
-.10	400		3.26	6.92	1.81	6.40	3.34
	∞		0	6.48	0	6.48	0
	200	.3863	4.27	7.42	3.10	6.14	4.23
0	400		3.46	7.78	2.21	6.31	3.44
	∞		0	8.38	0	8.38	0
	200	.3492	4.48	8.51	3.33	6.95	4.38
.03	400		3.26	8.99	2.02	7.08	3.22
	∞		0	12.79	0	12.79	0

monotonically increases as the magnitude of the $r_1(jk)$ correlation terms increase.

An inspection of Table 5.21, which presents the mean increase in actual error over optimum error, indicates that the behavior of the full, second and Matusita procedures exhibit greater stability across the various values of $r_1(jk)$ than either of the linear models. In fact, for fixed d_r with large $r_1(jk)$ correlation terms, the use of the first or LDF procedures resulted in significantly greater mean increases. Also, the table lends additional support to the contention that there are cases where discrimination for the linear procedures is improved by limiting the sample size, rather than by increasing it. (For example, both $\bar{\lambda}[1]$ and $\bar{\lambda}[\ell]$ increase with larger sample sizes whenever $d_r > 0$).

These last two experiments demonstrate that conclusions are, for the most part, generalizable across population structures containing identical d_r terms. Differences which do exist are of degree rather than direction. In fact, the evidence suggests that anomalies which result from the use of an inappropriate procedure are magnified as the correlations among the variables increase.

Further Monte Carlo sampling was accomplished on the correlation structures previously described in Table 5.16. However, in subsequent experiments the marginal probabilities were variable so that the effects of larger mean differences on performance could be examined. The following marginal probability vectors were considered:

$p_{1j}=(.2,.2,.2,.2,.2,.2)$, $p_{2j}=(.4,.4,.4,.4,.4,.4)$;

$p_{1j}=(.3,.3,.3,.3,.3,.3)$, $p_{2j}=(.6,.6,.6,.6,.6,.6)$;

$p_{1j}=(.2,.2,.2,.2,.2,.2)$, $p_{2j}=(.6,.6,.6,.6,.6,.6)$.

The results of the Monte Carlo sampling experiments for these population structures are summarized in Tables 5.22 through 5.24. A study of the tables yields the following conclusions:

1. Uniformly, for fixed d_r, both the optimum error and the theoretical errors decrease as d_p increases. Also, the use of the linear models yield theoretical errors which are essentially the same as the optimum error over a wider range of d_r values as d_p becomes large. Recall, identical results were obtained in prior experiments (e.g., Section 5.11.1) and it is again clear that with large mean differences most discrimination procedures do well.

2. For fixed d_r, the mean increase in actual error for each discrimination procedure decreases as the mean differences increase. This is not too surprising, particularly considering the behavior of the optimum error and the theoretical errors for large d_p.

TABLE 5.22

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.2,.3,.4
WITH ALL $r_1(jk)$=.10

d_r $[r_2(jk)-r_1(jk)]$	Optimum Error	Theoretical First	Theoretical LDF
p_{1j}=(.2,.2,.2,.2,.2,.2),p_{2j}=(.4,.4,.4,.4,.4,.4)			
-.15	.2693	.2695	.2695
-.10	.2890	.2890	.2890
-.05	.3084	.3084	.3084
0	.3278	.3278	.3278
.10	.3523	.3667	.3667
.20	.3306	.4056	.4056
p_{1j}=(.3,.3,.3,.3,.3,.3),p_{2j}=(.6,.6,.6,.6,.6,.6)			
-.15	.2236	.2236	.2236
-.10	.2406	.2406	.2406
-.05	.2579	.2579	.2579
0	.2752	.2752	.2752
.10	.2928	.3097	.3097
.20	.2828	.3443	.3443
p_{1j}=(.2,.2,.2,.2,.2,.2),p_{2j}=(.6,.6,.6,.6,.6,.6)			
-.15	.1527	.1527	.1527
-.10	.1698	.1698	.1698
-.05	.1870	.1870	.1870
0	.2043	.2043	.2043
.10	.2389	.2399	.2389
.20	.2402	.2734	.2739

TABLE 5.23

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR d_p=.2,.3, AND .4 WITH ALL $r_1(jk)$=.10

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Full	First	Second	LDF	Matusita
p_{1j}=(.2,.2,.2,.2,.2,.2), p_{2j}=(.4,.4,.4,.4,.4,.4)							
	200	.2693	5.46	1.54	2.24	2.32	5.00
-.15	400		2.97	0.85	0.84	1.63	3.00
	∞		0	0.02	0	0.02	0
	200	.2890	5.51	1.23	2.50	1.99	5.60
-.10	400		3.39	0.63	1.07	1.39	3.40
	∞		0	0	0	0	0
	200	.3084	5.59	0.79	2.79	1.45	5.69
-.05	400		3.44	0.36	1.45	1.05	3.48
	∞		0	0	0	0	0
	200	.3278	4.61	0.46	2.56	0.95	4.65
0	400		3.17	0.25	1.28	0.71	3.23
	∞		0	0	0	0	0
	200	.3523	3.07	1.18	1.93	1.03	3.10
.10	400		2.06	1.32	1.56	1.01	2.07
	∞		0	1.44	0	1.44	0
	200	.3306	3.24	6.51	2.00	5.40	3.19
.20	400		2.10	7.07	1.51	5.64	2.03
	∞		0	7.50	0	7.50	0
p_{1j}=(.3,.3,.3,.3,.3,.3), p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.2236	4.33	0.60	1.94	1.22	4.32
-.15	400		2.46	0.12	0.84	0.49	2.48
	∞		0	0	0	0	0
	200	.2406	4.19	0.35	2.06	1.03	4.20
-.10	400		2.70	0.11	1.16	0.54	2.73
	∞		0	0	0	0	0
	200	.2579	4.06	0.24	2.05	0.78	4.07
-.05	400		2.67	0.07	1.33	0.54	2.68
	∞		0	0	0	0	0
	200	.2752	3.52	0.14	2.08	0.55	3.52
0	400		2.44	0.03	1.34	0.24	2.46
	∞		0	0	0	0	0
	200	.2928	3.09	1.59	2.11	1.38	3.05
.10	400		2.30	1.66	1.62	1.40	2.29
	∞		0	1.69	0	1.69	0
	200	.2828	3.07	5.83	1.91	4.69	2.86
.20	400		2.01	6.07	1.15	5.05	1.91
	∞		0	6.15	0	6.15	0

TABLE 5.23 (CONTINUED)

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Full	First	Second	LDF	Matusita
p_{1j}=(.2,.2,.2,.2,.2,.2), p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.1527	2.97	0.06	1.14	0.21	2.97
-.15	400		1.55	0.01	0.72	0.02	1.56
	∞		0	0	0	0	0
	200	.1698	3.13	0.05	1.38	0.29	3.13
-.10	400		1.79	0.01	0.76	0.05	1.79
	∞		0	0	0	0	0
	200	.1870	3.01	0.04	1.51	0.29	3.01
-.05	400		1.75	0	0.81	0.07	1.76
	∞		0	0	0	0	0
	200	.2043	2.90	0.02	1.50	0.35	2.91
0	400		1.81	0	0.89	0.10	1.82
	∞		0	0	0	0	0
	200	.2389	1.98	0.02	1.16	0.10	1.95
.10	400		1.10	0	0.50	0.03	1.07
	∞		0	0	0	0	0
	200	.2402	2.68	3.30	2.13	2.88	2.57
	400		1.99	3.32	1.55	3.11	1.90
.20	∞		0	3.32	0	3.32	0

TABLE 5.24

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR dp=.2,.3,.4, AND ALL $r_1(jk)$=.10

d_r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.4,.4,.4,.4,.4,.4)							
	200	.3396	.3302	.8653	.3290	.4716	.084
-.15	400	.2898	.3356	.6951	.3340	.4000	.050
	∞	1.0	.3381	1.0	.3381	1.0	
	200	.3675	.9297	.8431	.9196	.4280	.001
-.10	400	.4218	.9341	.9047	.9238	.5088	-
	∞	1.0	.9398	1.0	.9398	1.0	
	200	.4305	.9722	.8678	.9508	.4557	-
-.05	400	.6141	.9739	.9100	.9644	.6305	-
	∞	1.0	.9822	1.0	.9822	1.0	
	200	.3615	.9395	.7452	.8670	.3831	.001
0	400	.5370	.9595	.8504	.9279	.5628	-
	∞	1.0	.9772	1.0	.9772	1.0	
	200	.4939	.9113	.7750	.8696	.6035	.040
.10	400	.6642	.9140	.9276	.8906	.7470	.015
	∞	1.0	.9202	1.0	.9202	1.0	
	200	.6806	.8248	.7072	.7805	.7485	.120
.20	400	.7706	.8238	.9323	.8106	.8474	.090
	∞	1.0	.8269	1.0	.8269	1.0	
p_{1j}=(.3,.3,.3,.3,.3,.3);p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.6181	.8421	.8433	.8435	.8051	.020
-.15	400	.7710	.8431	.9794	.8397	.8868	.012
	∞	1.0	.8461	1.0	.8461	1.0	
	200	.5931	.9678	.9206	.9564	.7226	.001
-.10	400	.7118	.9677	.9606	.9681	.7811	-
	∞	1.0	.9729	1.0	.9729	1.0	
	200	.5104	.9789	.8006	.9668	.6954	.002
-.05	400	.6967	.9831	.9513	.9741	.7903	-
	∞	1.0	.9885	1.0	.9885	1.0	
	200	.5463	.9621	.8804	.9121	.6621	.006
0	400	.6771	.9789	.9288	.9606	.8102	-
	∞	1.0	.9820	1.0	.9820	1.0	
	200	.3922	.8772	.6146	.8322	.6259	.012
.10	400	.5519	.8845	.9013	.8796	.7296	.001
	∞	1.0	.8864	1.0	.8864	1.0	
	200	.5455	.6106	.6786	.5617	.7520	.075
.20	400	.4101	.6162	.7450	.6027	.6661	.032
	∞	1.0	.6186	1.0	.6186	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

TABLE 5.24 (CONTINUED)

d_r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
$p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.6,.6,.6,.6,.6,.6)$							
	200	.7287	.8950	.9611	.8845	.8895	.022
-.15	400	.8035	.9043	.9567	.9028	.8907	.014
	∞	1.0	.9066	1.0	.9066	1.0	
	200	.7307	.9876	.9510	.9647	.7936	.007
-.10	400	.7716	.9885	.9724	.9834	.7654	-
	∞	1.0	.9903	1.0	.9903	1.0	
	200	.7144	.9837	.9551	.9649	.8087	.009
-.05	400	.8037	.9923	.9836	.9900	.8627	-
	∞	1.0	.9931	1.0	.9931	1.0	
	200	.5559	.9703	.8999	.9380	.7090	.007
0	400	.7013	.9821	.9598	.9701	.8265	-
	∞	1.0	.9838	1.0	.9838	1.0	
	200	.5227	.9217	.8319	.9054	.7317	.018
.10	400	.6613	.9236	.9429	.9069	.8207	.001
	∞	1.0	.9260	1.0	.9260	1.0	
	200	.5618	.7686	.6702	.7294	.7850	.085
.20	400	.7504	.7706	.7413	.7664	.9077	.032
	∞	1.0	.7711	1.0	.7711	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

3. With large $|d_r|$ and fixed d_p, the second and Matusita procedures yield larger mean correlations than any of the other models. However, generally speaking, the behavior of the mean correlations for each procedure is quite erratic across the various values of d_r, and, in fact, there are several cases where they decrease as both d_p and n increase.

5.12.1 - Conclusions--Sampling Under Case (ii): The population pairs considered under this case included correlation structures with $r_1(jk) \neq r_2(jk)$ for all $j \neq k$, $j,k=1,2,\ldots,6$ where all $r_1(jk)$ and $r_2(jk)$ were free to take on non-zero values.

The following summarizes the results of the series of Monte Carlo sampling experiments already described:

1. It was again possible to determine critical bounds on d_r, such that if they are exceeded the use of either of the linear models resulted in greater misclassification (on the average) than the full, second or Matusita procedures. Also, the observed L.L.R.'s for the first and LDF procedures were linearly different from the true L.L.R.'s whenever d_r assumed values beyond the critical bounds specified.

2. Anomalies that resulted when an inappropriate model was used were for the most part generalizable across different population structures for identical values of d_r. Differences which did exist were of degree rather than of direction; in fact, anomalies which can arise from the use

of inappropriate models are likely to be more severe when correlations are "high" rather than when "low".

3. For fixed d_r, the optimum error, theoretical error, and mean increase in actual error were all decreasing functions of d_p. With large mean differences, the performance of all the discrimination procedures was quite similar even for large positive or negative values of d_r.

<u>5.13 - Case (iii)</u>: For the population pairs considered under this case an additional restriction is imposed on the correlation structure in that the correlation matrices in the two populations are assumed to be identical: $r_1(jk)=r_2(jk)$ for all $j\neq k$, $j,k=1,2,\ldots,6$. Thus the correlation between any two variables in population i is identical to the correlation between those two variables in population 2. Recall, if $R_1=R_2$, then $S_1=S_2$ if and only if $p_{2j}=1-p_{1j}$ for $j=1,2,\ldots,m$. Therefore, it is possible to investigate population pairs where the covariances are equal, which is an underlying assumption used to derive the LDF, and cases where the covariances are not equal in the two populations. Since in the following experiments, $r_1(jk)=r_2(jk)$ for all j and k, the common value assigned to the correlation terms in both populations is denoted by r.

The next series of sampling experiments considered population pairs with parameter values $d_p=.1,.2,.3,.4$, $n=200$, 400 and $r=0,.1,.2,.3$, and .33. The values assigned to the individual marginal probabilities were identical to those used in the previous experiments. Summary results for these population structures are presented in Tables 5.25 through 5.27 and Figures 5.10 and 5.11.

Inspection of Table 5.25, which displays the optimum errors and theoretical errors for all values of d_p indicates that both optimum and theoretical errors decrease as d_p increases across all values of r. For fixed d_p, the use of either of the linear models results in theoretical errors which are essentially the same as the optimum error whenever $r \leq .1$. However, if $r > .1$, then across all values of d_p the theoretical errors for the first and LDF procedures are greater than the corresponding optimum error.

A similar conclusion is reached for the mean increase in actual error, displayed in Table 5.26. Uniformly, across all values of d_p, the linear models yield smaller mean increases in actual error than any of the other procedures if $r \leq .1$. However, if $r > .1$ then the linear models should not be used since they result in mean increases in actual error which are greater than those of the full, second or Matusita procedures.

A study of the behavior of the linear models at large values of r yields some rather surprising conclusions. Recall, in prior experiments whenever d_p was large, performance was relatively unaffected by the correlation structure even for large values of d_r and all of the discrimination procedures gave nearly identical results.

TABLE 5.25

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.1,.2,.3, and .4

r	Optimum Error	Theoretical First	Theoretical LDF	First $\alpha[1]-\alpha$	LDF $\alpha[\ell]-\alpha$
p_{1j}=(.4,.4,.4,.4,.4,.4);p_{2j}=(.5,.5,.5,.5,.5,.5)					
0	.3997	.3997	.3997	0	0
.1	.4232	.4232	.4232	0	0
.2	.4240	.4466	.4466	.0226	.0226
.3	.3863	.4700	.4700	.0837	.0837
.33	.3745	.4771	.4771	.1026	.1026
p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.4,.4,.4,.4,.4,.4)					
0	.2890	.2890	.2890	0	0
.1	.3278	.3278	.3278	0	0
.2	.3380	.3667	.3667	.0287	.0287
.3	.2623	.4056	.4056	.1433	.1433
.33	.2396	.4173	.4173	.1777	.1777
p_{1j}=(.3,.3,.3,.3,.3,.3);p_{2j}=(.6,.6,.6,.6,.6,.6)					
0	.2174	.2174	.2174	0	0
.1	.2752	.2752	.2752	0	0
.2	.2922	.3329	.3329	.0407	.0407
.3	.2120	.3670	.3670	.1550	.1550
.33	.1885	.4070	.4070	.2185	.2185
p_{1j}=(.2,.2,.2,.2,.2,.2); p_{2j}=(.6,.6,.6,.6,.6,.6)					
0	.1390	.1390	.1390	0	0
.1	.2043	.2043	.2043	0	0
.2	.2363	.2696	.2696	.0333	.0333
.3	.1786	.3349	.3349	.1563	.1563
.33	.1349	.3545	.3545	.2195	.2195

Table 5.26

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR d_p=.1,.2,.3, and .4

r	n	Optimum Error	Full	First	Second	LDF	Matusita
p_{1j}=(.4,.4,.4,.4,.4,.4);p_{2j}=(.5,.5,.5,.5,.5,.5)							
	200	.3997	4.86	1.19	2.97	1.30	4.87
0	400		3.62	0.67	1.77	0.74	3.65
	∞		0	0	0	0	0
	200	.4232	3.46	0.42	2.21	0.92	3.47
.1	400		2.63	0.18	1.46	0.44	2.67
	∞		0	0	0	0	0
	200	.4240	2.99	1.91	1.88	1.88	2.99
.2	400		2.11	2.06	1.31	1.74	2.12
	∞		0	2.26	0	2.26	0
	200	.3863	4.27	7.42	3.10	6.14	4.23
.3	400		3.46	7.76	2.21	6.31	3.44
	∞		0	8.37	0	8.37	0
	200	.3745	4.55	9.19	3.22	7.49	4.45
.33	400		3.33	9.70	1.95	7.60	3.25
	∞		0	10.26	0	10.26	0
p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.4,.4,.4,.4,.4,.4)							
	200	.2890	4.50	0.60	1.66	0.88	5.00
0	400		2.24	0.31	0.85	0.54	2.57
	∞		0	0	0	0	0
	200	.3278	4.61	0.46	2.56	0.95	4.93
.1	400		3.17	0.25	1.78	0.71	3.44
	∞		0	0	0	0	0
	200	.3380	2.97	3.17	1.69	3.31	3.04
.2	400		2.01	3.00	1.37	3.36	2.05
	∞		0	2.87	0	2.87	0
	200	.2623	3.71	14.48	1.83	13.04	3.75
.3	400		2.06	14.40	0.78	13.92	2.06
	∞		0	14.33	0	14.33	0
	200	.2396	3.23	17.82	1.24	15.80	3.26
.33	400		1.73	17.83	0.45	16.79	1.73
	∞		0	17.73	0	17.73	0

TABLE 5.26 (CONTINUED)

r	n	Optimum Error	Full	First	Second	LDF	Matusita
p_{1j}=(.3,.3,.3,.3,.3,.3);p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.2174	3.81	0.32	1.58	0.64	3.82
0	400		2.12	0.77	0.77	0.29	2.12
	∞		0	0	0	0	0
	200	.2752	3.52	0.14	2.08	0.55	3.52
.1	400		2.44	0.03	1.34	0.24	2.46
	∞		0	0	0	0	0
	200	.2922	2.62	4.00	1.92	3.49	2.71
.2	400		1.94	4.06	1.39	3.79	1.94
	∞		0	4.07	0	4.07	0
	200	.2120	4.06	17.69	2.42	15.05	4.01
.3	400		2.37	17.80	1.13	16.05	2.27
	∞		0	15.50	0	15.50	0
	200	.1885	3.34	21.60	2.16	17.10	3.25
.33	400		1.94	21.90	0.95	19.00	1.82
	∞		0	21.85	0	21.85	0
p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.1390	3.30	0.14	1.14	0.30	3.30
0	400		1.59	0.01	0.49	0.03	1.57
	∞		0	0	0	0	0
	200	.2043	2.90	0.02	1.50	0.35	2.91
.1	400		1.81	0	0.89	0.10	1.84
	∞		0	0	0	0	0
	200	.2363	3.20	3.25	2.38	2.76	3.25
.2	400		2.17	3.32	1.34	3.02	2.21
	∞		0	3.33	0	3.33	0
	200	.1786	3.58	15.47	2.42	13.05	3.58
.3	400		2.31	15.61	1.34	13.99	2.29
	∞		0	15.63	0	15.63	0
	200	.1349	3.63	21.70	2.39	18.46	3.62
.33	400		1.79	21.95	1.11	19.52	1.72
	∞		0	21.95	0	21.95	0

TABLE 5.27

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIO AND TRUE LOG LIKELIHOOD RATIO BASED ON 100 MONTE CARLO TRIALS FOR d_p=.1,.2,.3, AND .4

r		Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
p_{1j}=(.4,.4,.4,.4,.4,.4);p_{2j}=(.5,.5,.5,.5,.5,.5)							
	200	.4013	.9597	.6616	.9644	.4494	-
0	400	.3524	.9137	.7979	.9137	.3943	-
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.1677	.8625	.5666	.6939	.2897	.002
.1	400	.2547	.9661	.7187	.7441	.3574	-
	∞	1.0	.9789	1.0	.9789	1.0	
	200	.2586	.3890	.4052	.3990	.3010	.023
.2	400	.3306	.5950	.4563	.5223	.3973	.003
	∞	1.0	.6120	1.0	.6120	1.0	
	200	.4888	-.0257	.5618	-.0257	.4897	.125
.3	400	.5221	-.0257	.6492	-.0257	.5997	.104
	∞	1.0	-.0300	1.0	-.0300	1.0	
	200	.4694	-.0648	.6426	-.0403	.4364	.225
.33	400	.7126	-.0658	.6612	-.0480	.6715	.224
	∞	1.0	-.0684	1.0	-.0684	1.0	
p_{1j}=(.2,.2,.2,.2,.2,.2); p_{2j}=(.4,.4,.4,.4,.4,.4)							
	200	.5117	.9845	.7872	.9674	.4711	0.72
0	400	.6232	.9792	.8316	.9669	.6129	0.18
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.3615	.9395	.7452	.8670	.3831	.008
.1	400	.5370	.9595	.8504	.9279	.5628	-
	∞	1.0	.9772	1.0	.9772	1.0	
	200	.5169	.7896	.4523	.7146	.5859	.034
.2	400	.6627	.7908	.8761	.7585	.7384	.008
	∞	1.0	.7975	1.0	.7975	1.0	
	200	.5694	.4292	.5941	.4052	.6529	.141
.3	400	.4922	.4278	.5603	.3880	.6334	.075
	∞	1.0	.4315	1.0	.4315	1.0	
	200	.5724	.3039	.4856	.2377	.6676	.209
.33	400	.6675	.3068	.6331	.2807	.7603	.130
	∞	1.0	.3082	1.0	.3082	1.0	

[a]The average proportion of times π_1=(x;[2]) was less than zero.

TABLE 5.27 (CONTINUED)

r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
p_{1j}=(.3,.3,.3,.3,.3,.3);p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.6899	.9892	.8235	.9788	.7587	.008
0	400	.7927	.7898	.9634	.9898	.8883	-
		1.0	1.0	1.0	1.0	1.0	
	200	.5463	.9621	.8804	.9621	.6621	.006
.1	400	.6771	.9789	.9298	.9606	.8102	-
		1.0	.9820	1.0	.9820	1.0	
	200	.2069	.3345	.2667	.3197	.2217	.025
.2	400	.5131	.7081	.6686	.6948	.6842	.005
		1.0	.7101	1.0	.7101	1.0	
	200	.7390	-.0514	.8471	-.0486	.7593	.172
.3	400	.6409	-.0512	.7596	-.0478	.7195	.110
		1.0	-.0516	1.0	-.0516	1.0	
	200	.7631	-.3268	.7663	-.2928	.6514	.227
.33	400	.8512	-.3280	.7810	-.2927	.7733	.184
		1.0	-.3293	1.0	-.3293	1.0	
p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.6,.6,.6,.6,.6,.6)							
	200	.7320	.9925	.8735	.7743	.8155	.053
0	400	.8005	.9986	.9358	.9969	.7990	.018
		1.0	1.0	1.0	1.0	1.0	
	200	.5559	.9703	.8999	.9380	.7090	.008
.1	400	.7013	.9821	.9598	.9701	.8265	-
		1.0	.9838	1.0	.9838	1.0	
	200	.4742	.8158	.6015	.7897	.6808	.032
.2	400	.7866	.8156	.9190	.7945	.8851	.008
		1.0	.8174	1.0	.8174	1.0	
	200	.5835	.2998	.6094	.2894	.7441	.145
.3	400	.5920	.3001	.6642	.2861	.7540	.078
		1.0	.3006	1.0	.3006	1.0	
	200	.6551	.1060	.6705	.0952	.7747	.213
.33	400	.7082	.1066	.7624	.1036	.8220	.129
		1.0	.1068	1.0	.1068	1.0	

[a]The average proportion of times π_i=(x;[2]) was less than zero.

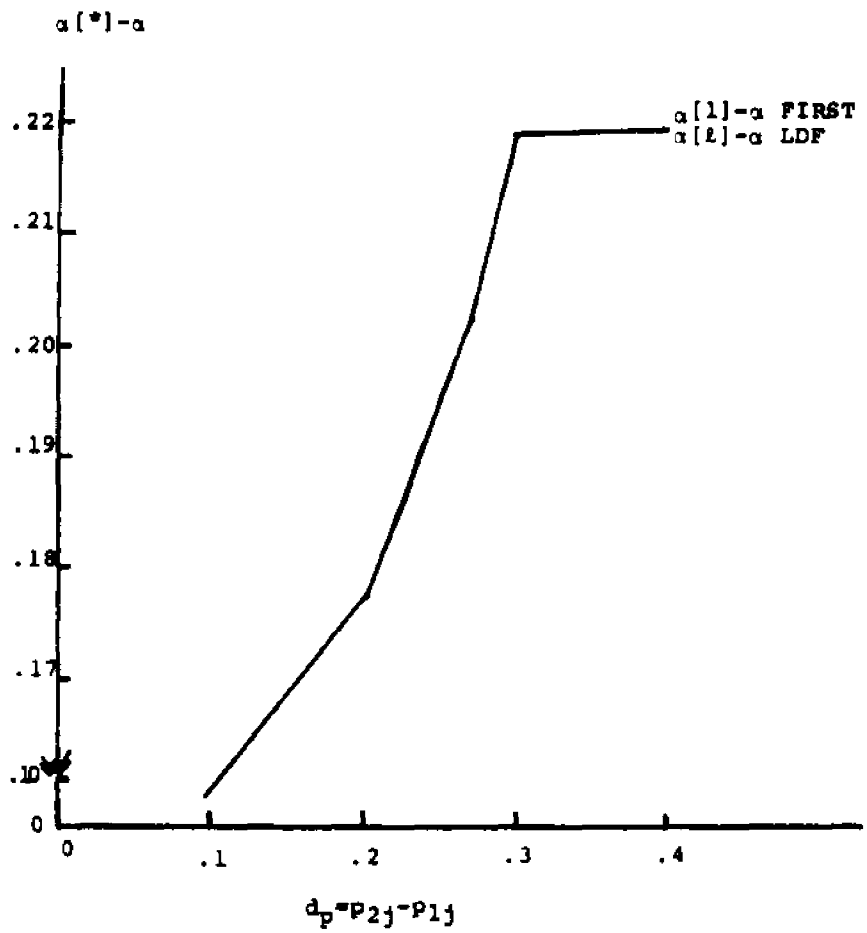

Figure 5.10 - Difference Between Theoretical Errors and Optimum Error for d_p=.1,.2,.3, and .4 at r=.33.

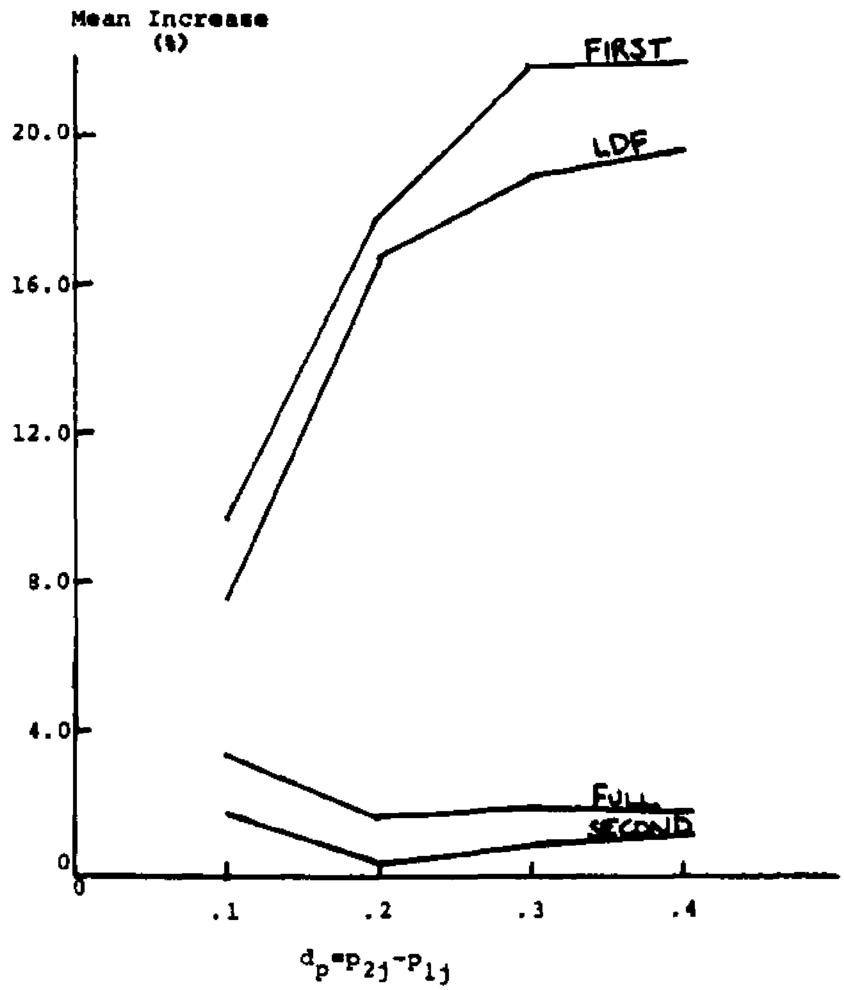

Figure 5.11 - Mean Increase in Actual Over Optimum Error (in Percent) Based on 100 Monte Carlo Trials for d_p=1,.2,.3, and .4 at r=.33.

With large d_p, the critical bounds on d_r such that the use of either of the linear models resulted in theoretical errors approximating the optimum error widened, and both $\alpha[1]-\alpha$, and $\alpha[\ell]-\alpha$ decreased. However, it is quite apparent that these conclusions do not hold for the population pairs now under examination.

Consider cases where r=.3 or .33. Here anomalies resulting from the use of the first or LDF procedures become more severe as d_p increases. For example, with r=.3 or .33 both $\alpha[1]-\alpha$, and $\alpha[\ell]-\alpha$ are increasing functions of d_p and, in fact, the theoretical errors are significantly greater than the corresponding optimum error. For example, at r=.33 and d_p=.4, α=.1349 while $\alpha[1]=[\ell]$=.3545. Similar irregularities occur for the mean increase in actual error over this range of r. Consider cases where r=.33 and n=400; here the mean increase in actual error increased from 9.70 percent, to 21.90 percent for the first, and from 7.60 percent to 19.00 percent for the LDF when d_p went from .1 to .4. Figures 5.10 and 5.11, which plot summary results for these two performance criteria, clearly illustrate the severity of using either of the linear models with "large" r.

The mean correlation between the observed L.L.R.'s and the true L.L.R.'s, displayed in Table 5.27, also substantiates the contention that even with large mean differences the performance of the linear models is significantly worse than the other procedures whenever $r \geq .3$. Uniformly, across all values of d_p and n, the use of the linear models at r=.3 or .33 yield significantly lower mean correlations than the other procedures and, in fact, there are several cases where both $\rho[1]$ and $\rho[\ell]$ are less than zero. It is also apparent that the Matusita procedure does particularly well for large values of r. At $r \geq .3.$, the mean correlations for this procedure are, for the most part, somewhat higher than either the full or second procedures. Note, however, the lower mean correlations for the second procedure may in part be due to the number of times the estimates $(\hat{\pi}_i(x;[2]))$ had to be set to 10^{-5}.

An explanation of these irregularities is deferred until after the next experiment which considers population pairs with identical variance-covariance structures. To accomplish this, values were assigned to p_{1j} and p_{2j} such that $p_{2j}=1-p_{1j}$ for j=1,2,....,6. In the following experiment $p_{1j}=.4$, $p_{2j}=.6$, j=1,2,....,6, n=200, 400 and r=0,.1.2,.3, and .33.

Table 5.28 and Figure 5.12 present summary results for the optimum error and the theoretical errors for the population pairs described above. The most striking point to be made from the table and figure is that even when the assumption of equal variance-covariance matrices is satisfied, the use of the linear models, and particularly the LDF, can still result in severe anomalies in the analysis.

TABLE 5.28

OPTIMUM AND THEORETICAL ERRORS FOR d_p=.2

WITH p_{1j}=.4, p_{2j}=.6, j=1,2,.....6

r	Optimum Error	Theoretical First	Theoretical LDF	First $\alpha[1]-\alpha$	LDF $\alpha[\ell]-\alpha$
0	.3174	.3174	.3174	-	-
.1	.3520	.3520	.3520	-	-
.2	.3578	.3866	.3866	.0288	.0288
.3	.2915	.4211	.4211	.1296	.1296
.33	.2716	.4315	.4315	.1599	.1599

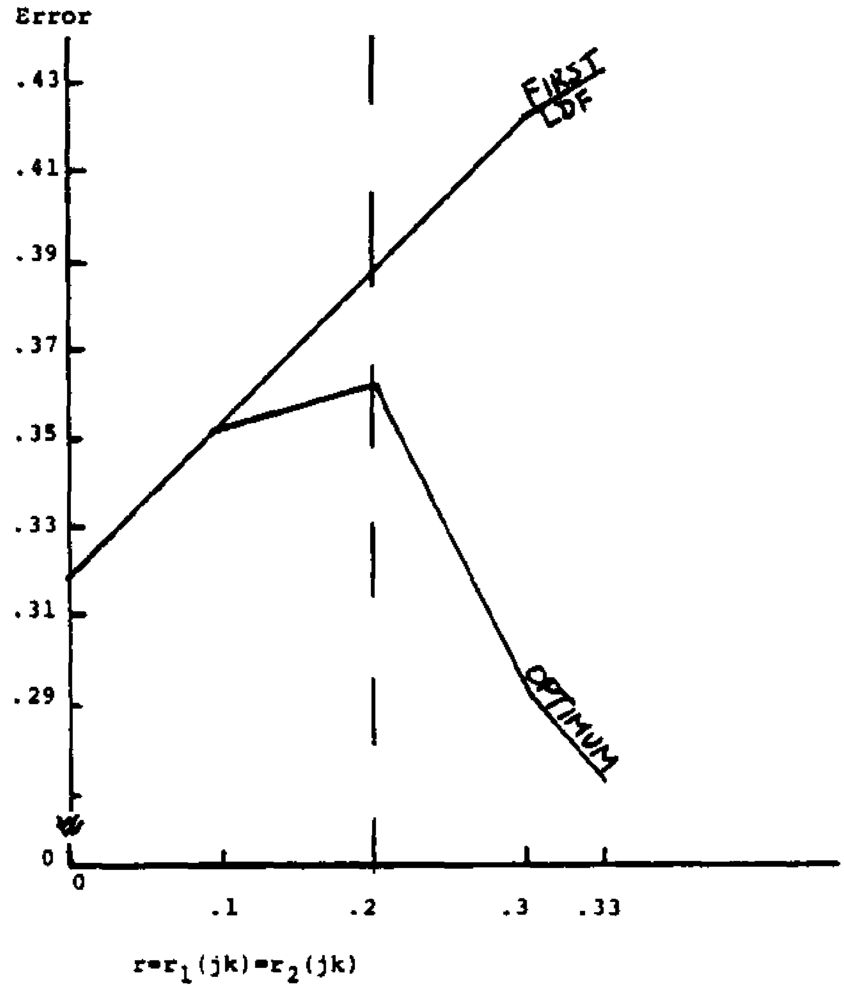

Figure 5.12 - Optimum and Theoretical Errors for d_p=.2 with p_{1j}=.4,p_{2j}=.6;j=1,2,......,6 and $r_1(jk)$=$r_2(jk)$=r.

It is obvious that the linear models yield theoretical errors which are significantly greater than the optimum error whenever $r>.2$. Note for all values of r, the first and LDF procedures have theoretical error which monotonically increase, while the optimum error at $r=.2$ reverses direction and continues to decrease for $r>.2$. Thus, correlated variables are again shown to discriminate better than uncorrelated variables.

A similar conclusion is reached for the mean increase in actual error displayed in Table 5.29. For $r\leq.1$, the use of the linear models resulted in the smallest mean increases for both sample sizes. At $r=.2$, the mean increase in actual error for the linear models increased somewhat but was not yet significantly different from the other procedures. However, at $r=.3$, a large jump in the order of magnitude of the mean increases occurred for both the first and LDF procedures. For example, at $n=400$, the mean increase in actual error increased from 2.88 percent to 12.96 percent for the first, and from 2.80 percent to 12.15 percent for the LDF when r went from .2 to .3. Also, the table indicates that as the sample size increased, both $\bar{\Lambda}[1]$ and $\bar{\Lambda}[\ell]$ increased at $r=.3$ and .33. This again implies that for these procedures, discrimination is improved by limiting the sample size, rather than by increasing it whenever $r\geq.3$.

An inspection of Table 5.30 which presents the mean correlations leads to the conclusion that there is a great

TABLE 5.29

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT)
BASED ON 100 MONTE CARLO TRIALS
FOR d_p=.2 WITH p_{1j}=.4,p_{2j}=.6;j=1,2,....6

r	n	Optimum Error	Full	First	Second	LDF	Matusita
	200	.3174	3.56	0.03	0.96	0.05	3.57
0	400		1.47	0	0.21	0	1.49
	∞		0	0	0	0	0
	200	.3520	3.04	0	1.81	0.21	3.04
.1	400		2.88	0	1.11	0.02	2.10
	∞		0	0	0	0	0
	200	.3578	2.60	2.88	1.32	2.62	2.61
.2	400		1.58	2.88	0.97	2.80	1.59
	∞		0	2.88	0	2.88	0
	200	.2195	4.06	12.96	1.61	11.19	4.03
.3	400		1.91	12.96	0.41	12.15	1.91
	∞		0	12.96	0	12.96	0
	200	.2716	3.52	15.96	1.10	13.44	3.56
.33	400		1.43	15.98	0.18	14.64	1.45
	∞		0	15.99	0	15.99	0

TABLE 5.30

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR dp=.2 WITH p_{1j}=.4, p_{2j}=.6;j=1,2,....6

r	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
	200	.4521	.9847	.9116	.9843	.5919	.001
0	400	.6761	.9832	.9393	.9773	.7196	-
	∞	1.0	1.0	1.0	1.0	1.0	
	200	.3388	.9704	.8928	.9509	.5293	.002
.1	400	.6415	.9717	.8859	.9534	.7555	-
	∞	1.0	.9806	1.0	.9806	1.0	
	200	.0415	.6408	.3542	.5644	.1916	.022
.2	400	.2938	.6534	.7246	.6196	.4539	-
	∞	1.0	.6580	1.0	.6580	1.0	
	200	.4903	-.0529	.3588	-.0411	.5689	.144
.3	400	.5330	-.0545	.5632	-.0492	.6679	.057
	∞	1.0	-.0550	1.0	-.0550	1.0	
	200	.6614	-.1963	.6145	-.1542	.7468	.238
.33	400	.5857	-.1967	.5939	-.1767	.6983	.069
	∞	1.0	-.2011	1.0	-.2011	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

deal of disparity between the observed L.L.R.'s and the true L.L.R.'s for the linear models even when the true state probabilities are assumed known. In fact, for $r>.3$, both $\alpha[1]$ and $\alpha[\ell]$ are less than zero which implies that the use of either of these models will result in greater misclassification error (on the average) than any of the other procedures.

Thus, the sampling experiments of this section have shown that there is a value of r, say r_o, such that if $r>r_o$, the use of the linear models will result in significantly greater errors than the other procedures. It is most important to note that even when the samples were taken from populations having the same variance-covariance structure, which is a basic underlying assumption used to derive the LDF, essentially the same anomalies in the analysis occurred. It should be mentioned that the last experiment was repeated except that in this example $d_p=.6$ with $p_{1j}=.2$ and $p_{2j}=.8$, for $j=1,2,....,6$. Here the decision was not to specifically show the summary tables, since results were nearly identical to those derived in the previous experiment. However, suffice to say the results again clearly demonstrate that significantly greater misclassification errors (on the average) can be expected even when the variance-covariance matrices are identical if the linear models are applied to population structures containing "highly" correlated variables.

The rather irregular behavior of the linear models for large r can be explained by examining the true L.L.R.'s as follows:

Consider population pairs with $p_{1j}=.4$ and $p_{2j}=.6$, $j=1,2,\ldots,6$. The log likelihood ratios for each discrimination procedure at $r=.3$ are shown graphically in Figure 5.13. These ratios were formed by substituting the true p_{ij} and $r_i(jk)$ into the formulas of Section 4.2. The L.L.R.'s can be plotted with respect to the number of positive X_j in the response vector x since all p_{ij} and all $r_i(jk)$ are identical in each of the two populations.

The figure reveals that the L.L.R.'s for the full, second and Matusita procedures do not increase monotonically with the number of positive X_j. The L.L.R.'s for these procedures are initially negative, become positive, then reverse both direction and sign at $\Sigma X_j=4$ before becoming positive again at $\Sigma X_j=5,6$. Moore (1973), who first demonstrated the possible non-monotonic behavior of the true L.L.R.'s, describes this non-monotonicity by stating that the L.L.R.'s undergo a "reversal" (p.402). It is important to note that this example illustrates a limitation of all linear models since it will always be impossible for these procedures to follow the true L.L.R.'s whenever they are non-monotonic.

<u>5.13.1 - Conclusions: Sampling Under Case (iii)</u>: A number of rather interesting and important conclusions were derived for population structures considered under this case. The population pairs included in this section contained correlation structures such that the correlation between any

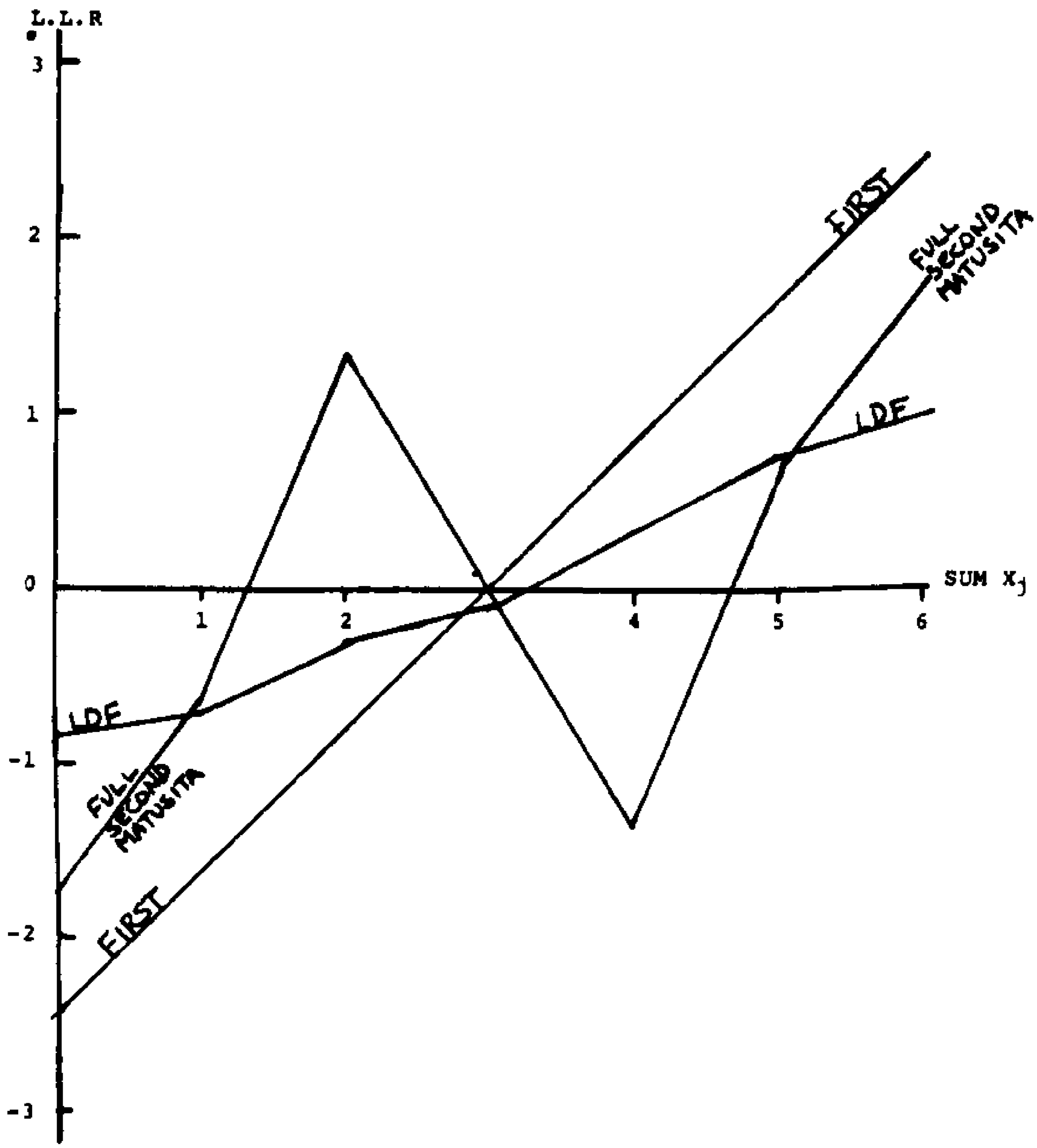

Figure 5.13 - Log Likelihood Ratios for d_p=.2 with p_{1j}=.4, p_{2j}=.6; j=1,2,.....,6 at r=.30.

two variables in population 1 was identical to the correlation between those two variables in population 2, i.e., $R_1=R_2$. In this manner, it was possible to investigate population pairs where the covariances were not equal in the two populations.

Among the most interesting of the findings were the following:

1. It was again possible to determine critical value of r, say r_o, such that if $r>r_o$, the use of the linear procedures resulted in significantly greater misclassification error (on the average) than any of the other procedures. It was also shown that if $r>r_o$, then discrimination with the linear procedures could be improved by limiting the sample size rather than by increasing it.

2. For values of r exceeding the critical limit r_o, the performance of the linear models did not improve with larger mean differences. Recall, in previous sampling experiments the linear models yielded essentially the same results as the other procedures as long as d_p was large, even with large $|d_r|$. However, here this was clearly not the case. Rather, for these population pairs, anomalies which surfaced with small mean differences were magnified as the difference between the mean vectors increased whenever $r>r_o$. Note the values of r such that $r>r_o$ correspond to the points at which reversals were shown to occur.

3. Perhaps the single most important fact revealed in this section was that severe anomalies in the analysis can occur even when the variance-covariance matrices in the two populations are taken to be identical. The true L.L.R.'s were shown to be capable of reversing both direction and sign and, hence, the linear models cannot satisfactorily characterize populations of this kind since their L.L.R.'s are always monotone.

5.2 - QUESTION 2 RESULTS

> If the difference between the mean vectors is fixed, but the magnitude of the individual p_{ij} is allowed to vary, what are the effects on the optimum error and the theoretical errors?

The previous experiments have shown that the effects on performance of increasing the difference between the marginal probability vectors in the two populations were not uniform across different populations. Rather, results were a function of the parameters $r_1(jk)$ and $r_2(jk)$. For cases where $R_1 \neq R_2$, the performance of each discrimination was improved as the mean differences increased. However, in populations where reversals were shown to occur, structures for which $R_1 = R_2$, anomalies uniformly surfaced for both small and large values of d_p. In fact, for these cases the further apart the underlying distributions, with respect to mean differences, the more severe were the irregularities.

Given these results, it was thought to be of interest to determine whether, for a fixed value of d_p, discrimination is also a function of the magnitude of the marginal probabilities. To accomplish this, several of the previous sampling experiments were replicated in the sense that the parameter values for d_p and $r_i(jk)$ did not change. However, in the replicated experiments a new set of marginal probabilities was used. In the following experiments, let p^*_{ij} denote the new set of marginal probabilities where in all cases $p^*_{ij} \neq p_{ij}$ and $p^*_{2j}-p^*_{1j}=d_p=p_{2j}-p_{1j}$, for $i=1,2,j=1,2,\ldots,6$. Initially, two examples were chosen from structures characterized by Case (i). For these examples the value of d_p was first set to .2, and then to .4. The marginal probability vectors corresponding to each value of d_p were:

$$\left.\begin{array}{l} p_{1j}=(.2,.2,.2,.2,.2,.2)\ ,\ p_{2j}=(.4,.4,.4,.4,.4,.4), \\ p^*_{1j}=(.3,.3,.3,.3,.3,.3)\ ,\ p^*_{2j}=(.5,.5,.5,.5,.5,.5); \end{array}\right. \quad d_p=.2$$

and

$$\left.\begin{array}{l} p^*_{1j}=(.1,.1,.1,.1,.1,.1)\ ,\ p^*_{2j}=(.5,.5,.5,.5,.5,.5), \\ p_{1j}=(.2,.2,.2,.2,.2,.2)\ ,\ p_{2j}=(.6,.6,.6,.6,.6,.6) \end{array}\right. \quad d_p=.4$$

Hereafter for notational convenience, the smaller of the two sets of p_{ij} values for each value of d_p will be referred to as Condition A, while the larger of the two sets will be referred to as Condition B. Summary results for $d_p=.2$ are displayed in Table 5.31 and Figure 5.14.

It is apparent from the table that the choice of a particular discrimination procedure does not appreciably

TABLE 5.31

OPTIMUM AND THEORETICAL ERRORS FOR VARIOUS VALUES OF p_{ij} WITH d_p=.2

d_r	Optimum Error		Theoretical			
			First		LDF	
$[r_2(13)]$	Condition A	Condition B	Condition A	Condition B	Condition A	Condition B
-.6	.2487	.2647	.2734	.2903	.2616	.2980
-.5	.2614	.2788	.2760	.2919	.2683	.2988
-.4	.2710	.2944	.2786	.2935	.2786	.2935
-.3	.2787	.2950	.2812	.2950	.2812	.2950
-.2	.2828	.2966	.2838	.2966	.2838	.2966
-.1	.2864	.2982	.2864	.2982	.2864	.2982
0	.2890	.2997	.2890	.2997	.2890	.2997
.1	.2916	.3013	.2916	.3013	.2916	.3013
.2	.2941	.3028	.2941	.3028	.2941	.3028
.3	.2893	.3044	.2967	.3044	.3036	.3044
.4	.2836	.3050	.2993	.3060	.3050	.3060
.5	.2779	.2924	.3019	.3075	.3060	.3075
.6	.2722	.2792	.3045	.3091	.3080	.3091

Condition A: p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.4,.4,.4,.4,.4,.4)

Condition B: p_{1j}=(.3,.3,.3,.3,.3,.3);p_{2j}=(.5,.5,.5,.5,.5,.5)

Condition A: $p_{1j}=(.2,.2,.2,.2,.2,.2)$, $p_{2j}=(.4,.4,.4,.4,.4,.4)$
Condition B: $p_{1j}=(.3,.3,.3,.3,.3,.3)$, $p_{2j}=(.5,.5,.5,.5,.5,.5)$

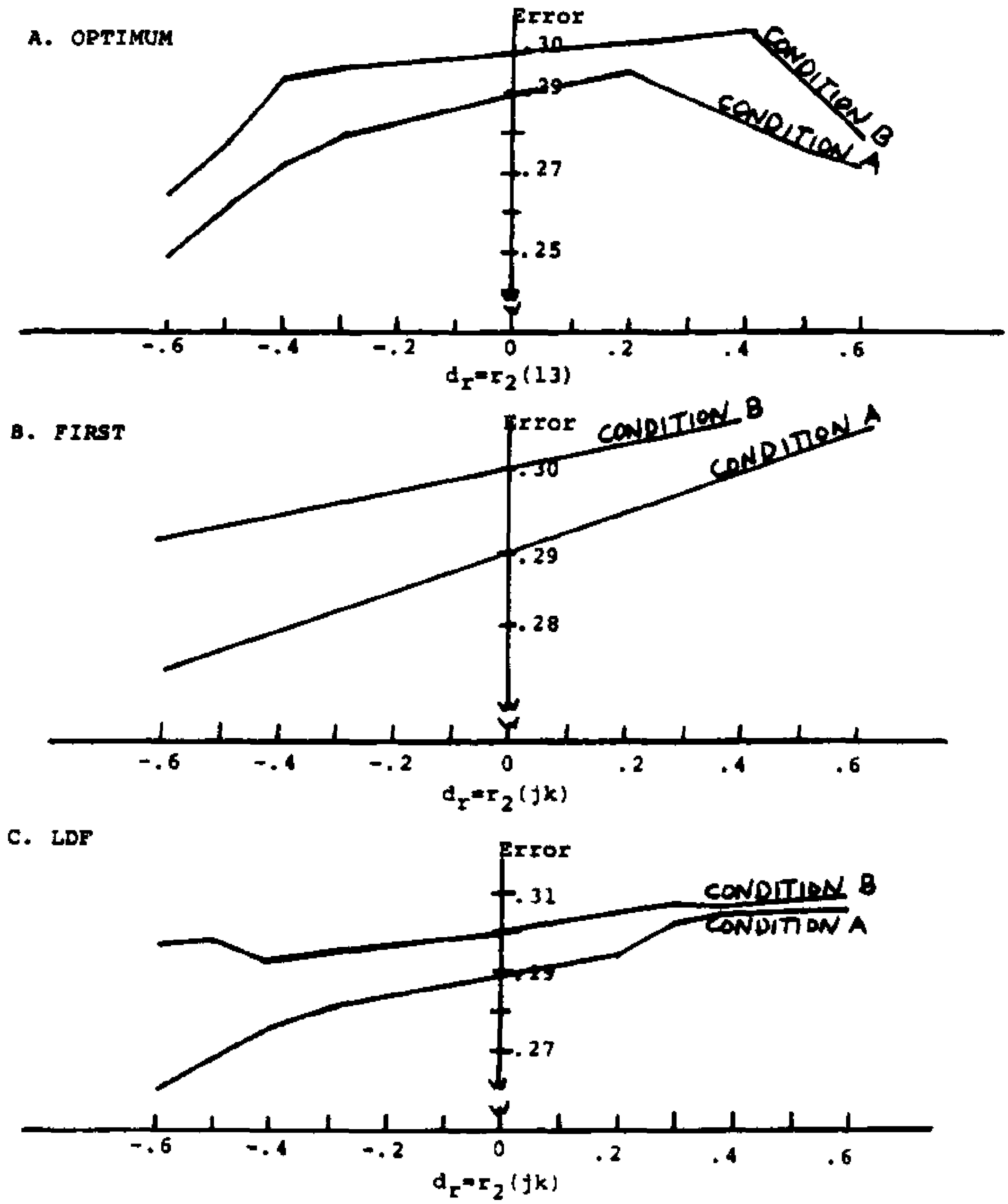

Figure 5.14 - Optimum and Theoretical Errors for Various Values of p_{ij} with $d_p=.2$.

change as the magnitude of the marginal probabilities is increased (Condition B). Uniformly, the behavior of both the optimum error and the theoretical errors under Condition B parrots that of Condition A as d_r is varied. Also, for fixed d_r, the difference between the theoretical errors and the optimum error, which can be easily derived from the table, are quite similar across both conditions.

A study of Figure 5.14 leads to the conclusion that the error rates associated with Condition B, the larger of the two sets of marginal probability vectors, are uniformly larger than those obtained under Condition A for all values of d_r. The equation for the optimum error and the theoretical errors as a function of d_r under Condition B is shown to lie above that for Condition A. Hence, the evidence suggests that discrimination becomes more difficult as the magnitude of the marginal probabilities is increased uniformly in both populations.

Summary results for population pairs with d_p=.4 are presented in Table 5.32 and Figure 5.15. The results for these examples are quite similar to those derived in the previous experiment. For fixed d_r both Condition A and Condition B lead to identical recommendations for selecting a particular discrimination procedure. In addition, the error rates under Condition B are again uniformly larger than those obtained under Condition A across all values of d_r. Note the error rates for the first procedure

TABLE 5.32

OPTIMUM AND THEORETICAL ERRORS FOR VARIOUS VALUES OF p_{ij} WITH d_p=.4

d_r [r_2(13)]	Optimum Error		Theoretical			
			First		LDF	
	Condition A	Condition B	Condition A	Condition B	Condition A	Condition B
-.6	.0967	.1198	.0978	.1252	.0968	.1306
-.5	.1001	.1262	.1001	.1275	.1039	.1326
-.4	.1024	.1298	.1024	.1298	.1109	.1298
-.3	.1048	.1321	.1048	.1321	.1048	.1321
-.2	.1071	.1344	.1071	.1344	.1071	.1344
-.1	.1095	.1367	.1095	.1367	.1095	.1367
0	.1111	.1390	.1118	.1390	.1118	.1390
.1	.1142	.1413	.1142	.1413	.1142	.1413
.2	.1165	.1436	.1165	.1436	.1165	.1436
.3	.1189	.1460	.1189	.1460	.1257	.1460
.4	.1212	.1483	.1212	.1483	.1289	.1483
.5	.1235	.1506	.1235	.1506	.1320	.1506
.6	.1246	.1495	.1292	.1520	.1351	.1529

Condition A: p_{1j}=(.1,.1,.1,.1,.1,.1), p_{2j}=(.5,.5,.5,.5,.5,.5)

Condition B: p_{1j}=(.2,.2,.2,.2,.2,.2), p_{2j}=(.6,.6,.6,.6,.6,.6)

Condition A: p_{1j}=(.1,.1,.1,.1,.1,.1), p_{2j}=(.5,.5,.5,.5,.5,.5)

Condition B: p_{1j}=(.2,.2,.2,.2,.2,.2), p_{2j}=(.6,.6,.6,.6,.6,.6)

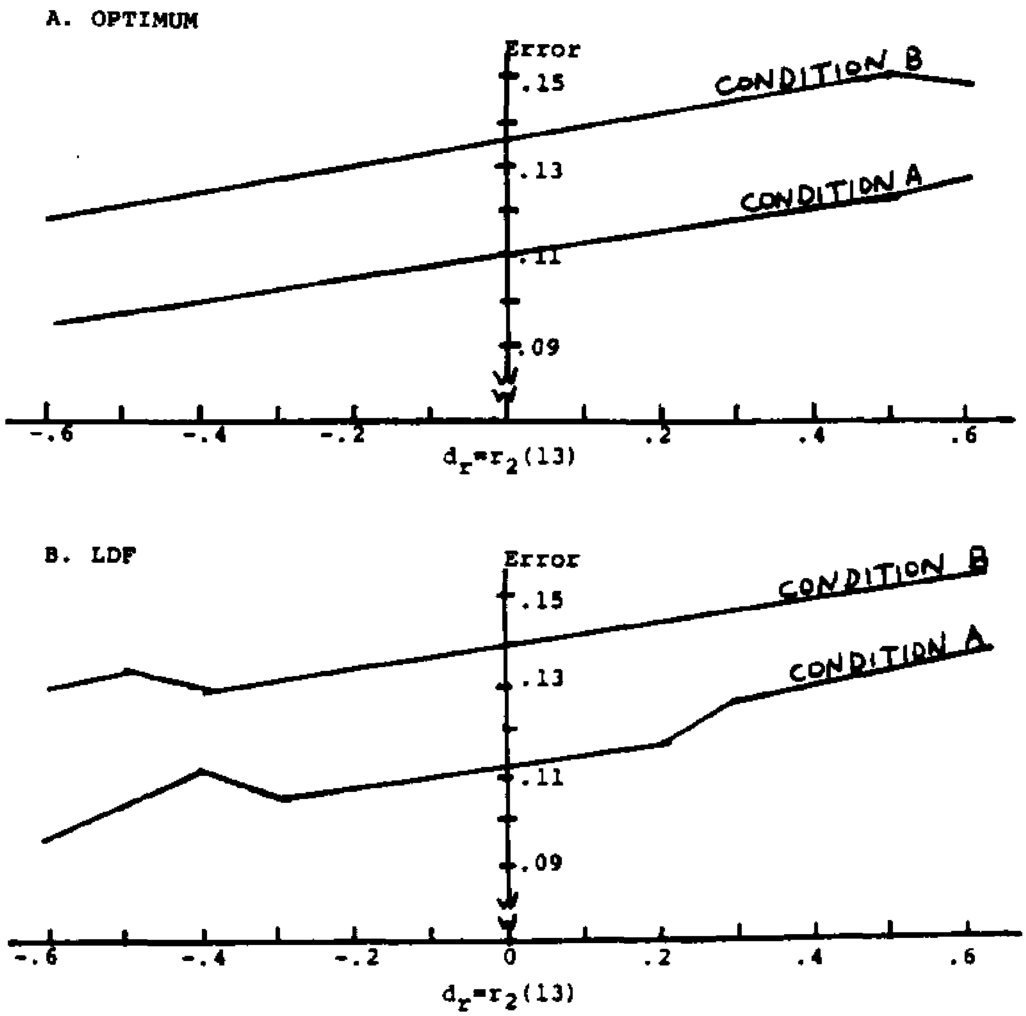

Figure 5.15 - Optimum and Theoretical Errors for Various Values of p_{ij} with d_p=.4.

Condition A: p_{1j}=(.1,.1,.1,.1,.1,.1),p_{2j}=(.5,.5,.5,.5,.5,.5)

Condition B: p_{1j}=(.2,.2,.2,.2,.2,.2),p_{2j}=(.6,.6,.6,.6,.6,.6)

C. FIRST

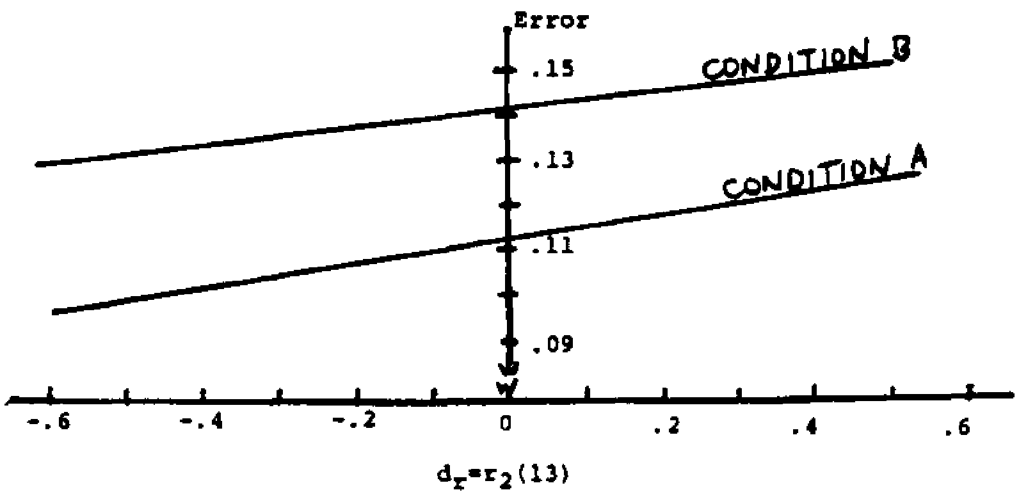

Figure 5.15 (continued)

shift up parallel as the magnitude of the marginal probabilities is increased.

The next series of experiments considered two representative examples from Case (iii). In these experiments, d_p was first set to .1 and then to .3 where the corresponding marginal probability vectors were taken to be of the form:

$$\left.\begin{array}{l} p^*_{1j}=(.2,.2,.2,.2,.2,.2)\ ,\ p^*_{2j}=(.3,.3,.3,.3,.3,.3), \\ p_{1j}=(.4,.4,.4,.4,.4,.4)\ ,\ p_{2j}=(.5,.5,.5,.5,.5,.5); \end{array}\right. \quad d_p=.1$$

and

$$\left.\begin{array}{l} p^*_{1j}=(.2,.2,.2,.2,.2,.2)\ ,\ p^*_{2j}=(.5,.5,.5,.5,.5,.5), \\ p_{1j}=(.3,.3,.3,.3,.3,,3)\ ,\ p_{2j}=(.6,16,16,16,16,16). \end{array}\right. \quad d_p=.3$$

Tables 5.33 and 5.34 display the optimum error and the theoretical errors for these examples.

A study of the tables indicates that the error rates obtained under Condition A and Condition B are rather close and hence, for fixed r the relative performance of each discrimination procedure does not appreciably change. However, in these experiments the behavior of the error rates under the two conditions seems somewhat more erratic than in the previous experiments. Recall, in previous experiments the optimum error and the theoretical errors were uniformly lower under Condition A than under Condition B, for fixed d_p and d_r. Although this relationship again holds for the

TABLE 5.33

OPTIMUM AND THEORETICAL ERRORS FOR VARIOUS VALUES OF p_{ij} WITH dp=.1

	Optimum Error		Theoretical			
			First		LDF	
r	Condition A	Condition B	Condition A	Condition B	Condition A	Condition B
0	.3752	.3997	.3752	.3997	.3752	.3997
.1	.4094	.4232	.4094	.4232	.4094	.4232
.2	.4051	.4240	.4364	.4466	.4364	.4466
.3	.3603	.3863	.4634	.4700	.4634	.4700
.33	.3840	.3745	.4968	.4771	.4968	.4771

Condition A: p_{1j}=(.2,.2,.2,.2,.2,.2);p_{2j}=(.3,.3,.3,.3,.3,.3)

Condition B: p_{1j}=(.4,.4,.4,.4,.4,.4);p_{2j}=(.5,.5,.5,.5,.5,.5)

TABLE 5.34

OPTIMUM AND THEORETICAL ERRORS FOR VARIOUS VALUES OF p_{ij} WITH d_p=.3

r	Optimum Error Condition A	Optimum Error Condition B	Theoretical First Condition A	Theoretical First Condition B	Theoretical LDF Condition A	Theoretical LDF Condition B
0	.2061	.2174	.2061	.2174	.2061	.2174
.1	.2622	.2752	.2755	.2752	.2755	.2752
.2	.2881	.2922	.3296	.3329	.3296	.3329
.3	.1930	.2120	.3830	.3670	.3830	.3670
.33	.1631	.1885	.4000	.4070	.4000	.4070

Condition A: p_{1j}=(.2,.2,.2,.2,.2,.2); p_{2j}=(.5,.5,.5,.5,.5,.5)

Condition B: p_{1j}=(.3,.3,.3,.3,.3,.3); p_{2j}=(.6,.6,.6,.6,.6,.6) optimum error, the theoretical errors behave somewhat differently. For example, if $r \geq .3$ then the theoretical errors under Condition A are uniformly greater than those obtained under Condition B. Note this irregularity occurs only for the theoretical errors (linear models) at $r=.3$ or .33 and hence it may in part be due to the non-monotonicity of the L.L.R.'s since it is for these values of r that reversals were shown to exist.

5.21 - Conclusions - Question 2: The following summarizes the results of the sampling experiments previously described.

1. Uniformly, across all values of d_p, the optimum error and the theoretical errors under Condition A and Condition B were rather close and, hence, for fixed d_r the relative performance of each discrimination procedure did not appreciably change. Thus, it would seem that the selection of a particular procedure is more a function of the values of d_p and d_r (or r) rather than the parameters p_{1j} and p_{2j}.

2. Although the relative performance of each discrimination procedure did not appreciably change as the magnitude of the marginal probabilities was varied, the magnitude of the resultant error rates, however, did. Larger optimum errors were obtained across all population structures when the magnitude of the individual marginal probabilities increased. Thus, the evidence suggests that

for a given value of d_p, discrimination becomes more difficult as the magnitude of the marginal probabilities increases. A similar relationship held for the theoretical errors except in those population structures where reversals occurred. Here results were less exact.

5.3 - QUESTION 3 - RESULTS

> Traditionally, the discrimination problem has been addressed in terms of differences in mean structure. However, it would be interesting to determine whether certain classification procedures can effectively discriminate on the basis of correlation structure rather than on mean differences.

The following experiments were designed to investigate whether discrimination can be initiated on population pairs containing similar mean vectors but which vary with respect to their correlation structure. The ability to discriminate on the basis of the correlations among variables seems particularly valuable since often the marketing researcher must handle data in which there exists only a limited number of variables having significantly different mean scores across populations. Thus, it is not surprising to note that under these circumstances the use of a discrimination procedure such as the Fisher LDF, which is based on mean differences, yields error rates which are not significantly different from chance assignment. However, the problem of insignificant mean differences may be negotiated by the use of alternative classification procedures which are less sensitive to mean scores and which use the information provided by the correlations among variables.

Monte Carlo sampling experiments were accomplished for 20 pairs of populations with mean differences between 0 and .10. These population pairs are described in Table 5.35, and can be divided into five groups. Group I consists of the population pairs with all correlations $r_i(jk)=.10$(pairs 1, 6, 11, 16). For the remaining groups the correlation structures were assumed to be different in the two populations. For example, Group II contained pairs for which $r_1(jk)=.10$, and $r_2(jk)=.30$ (pairs, 2, 7, 12, 17). Group III consists of those pairs with $r_1(jk)=.20$ and $r_2(jk)=.05$ (pairs 3, 8, 13, 18). Group IV includes those pairs for which $r_1(jk)=.30$, and $r_2(jk)=0$ (pairs 4, 9, 14, 19). Finally, Group V contained pairs with $r_1(jk)=.33$, and $r_2(jk)=-.05$ (pairs 5, 10, 15, 20).

The results of the Monte Carlo sampling are summarized in Tables 5.36 through 5.38. Since there existed a great deal of similarity among sample population pairs within the same group, the pairs displayed in the tables were chosen as representative. The results for Group I are represented by pairs 6 and 11, while pairs 7, 12, and 17 are given for Group II. Pairs 3, 8, 13 and 18, and pairs 4, 9, 14 and 19 are used to summarize results from Group III and Group IV, respectively. Results for Group V are represented by pairs 5, 10, 15 and 20. The third row for each population pair in Table 5.36 indicates whether the resultant mean actual non-error rate for each procedure did better than chance (denoted in the table by $n=\infty$). To

TABLE 5.35

SAMPLE POPULATION PAIRS

Pair	Group	Population 1 p_{1j}	$r_1(jk)$	Population 2 p_{2j}	$r_2(jk)$
1	I	(.5,.5,.5,.5,.5,.5)	All $r_1(jk)=.10$	(.5,.5,.5,.5,.5,.5)	All $r_2(jk)=.10$
2	II	"	All $r_1(jk)=.10$	"	All $r_2(jk)=.30$
3	III	"	All $r_1(jk)=.20$	"	All $r_2(jk)=.05$
4	IV	"	All $r_1(jk)=.30$	"	All $r_2(jk)=0$
5	V	"	All $r_1(jk)=.33$	"	All $r_2(jk)=-.05$
6	I	(.48,.48,.48,.48, .48,.48)	All $r_1(jk)=.10$	(.5,.5,.5,.5,.5,.5)	All $r_2(jk)=.10$
7	II	"	All $r_1(jk)=.10$	"	All $r_2(jk)=.30$
8	III	"	All $r_1(jk)=.20$	"	All $r_2(jk)=.05$
9	IV	"	All $r_1(jk)=.30$	"	All $r_2(jk)=0$
10	V	"	All $r_1(jk)=.33$		All $r_2(jk)=-.05$
11	I	(.42,.42,.42,.42, .42,.42)	All $r_1(jk)=.10$	(.5,.5,.5,.5,.5,.5)	All $r_2(jk)=.10$
12	II	"	All $r_1(jk)=.10$	"	All $r_2(jk)=.30$
13	III	"	All $r_1(jk)=.20$	"	All $r_2(jk)=.05$
14	IV	"	All $r_1(jk)=.30$	"	All $r_2(jk)=0$
15	V	"	All $r_1(jk)=.33$	"	All $r_2(jk)=-.05$
16	I	(.4,.4,.4,.4,.4,.4)	All $r_1(jk)=.10$	(.5,.5,.5,.5,.5,.5)	All $r_2(jk)=.10$
17	II	"	All $r_1(jk)=.10$	"	All $r_2(jk)=.30$
18	III	"	All $r_1(jk)=.20$	"	All $r_2(jk)=.05$
19	IV	"	All $r_1(jk)=.30$	"	All $r_2(jk)=0$
20	V	"	All $r_1(jk)=.33$	"	All $r_2(jk)=-.05$

TABLE 5.36

OPTIMUM AND THEORETICAL ERRORS FOR POPULATION PAIRS WITH $d_p \leq .1$

Pair	Group	Optimum Error	Theoretical First	Theoretical LDF	First $\alpha[1]-\alpha$	LDF $\alpha[\ell]-\alpha$
6	I	.4849	.4849	.4849	-	-
11	I	.4387	.4387	.4387	-	-
7	II	.3602	.5318	.5318	.1716	.1716
12	II	.3718	.4856	.4856	.1138	.1138
17	II	.3736	.4700	.4700	.0964	.0964
3	III	.3945	.5123	.5211	.1178	.1266
8	III	.3941	.4537	.4537	.0596	.0596
13	III	.3874	.4218	.4218	.0344	.0344
18	III	.3779	.4114	.4114	.0335	.0335
4	IV	.2891	.4684	.4578	.1793	.1687
9	IV	.2889	.4226	.4226	.1337	.1337
14	IV	.2874	.4049	.4049	.1175	.1175
19	IV	.2865	.3997	.3997	.1132	.1132
5	V	.2328	.5579	.5534	.3251	.3251
10	V	.2328	.4050	.4050	.1722	.1722
15	V	.2327	.3917	.3917	.1590	.1590
20	V	.2328	.3880	.3880	.1552	.1552

TABLE 5.37

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR POPULATION PAIRS WITH $d_p \leq .1$

Pair	Group	Optimum	n	Full	First	Second	LDF	Matusita
			200	1.28	0.92	1.18	1.03	1.28
6	I	.4899	400	1.13	0.40	0.91	0.60	1.14
			∞					
			200	3.17	0.48	2.06	1.05	3.18
11	I	.4387	400	2.36	0.18	1.41	0.45	2.39
			∞					
			200	4.62	13.52	2.59	13.57	4.57
7	II	.3602	400	2.89	13.29	1.37	13.41	2.87
			∞	*		**		*
			200	3.63	8.63	2.11	8.78	3.51
12	II	.3718	400	2.26	8.81	1.31	8.45	2.18
			∞	*		**		*
			200	3.22	7.47	1.91	7.20	3.11
17	II	.3736	400	2.10	7.79	1.18	7.06	2.02
			∞	*		**		*
			200	4.82	10.55	2.71	10.60	4.86
3	III	.3945	400	3.52	10.55	1.53	10.54	3.61
			∞	*		*		*
			200	4.78	9.98	2.83	10.14	4.86
8	III	.3941	400	3.48	9.70	1.62	9.85	3.57
			∞	*		*		*
			200	4.46	5.68	2.88	6.32	4.49
13	III	.3874	400	3.15	5.33	1.70	5.72	3.30
			∞	*		*		*
			200	4.52	5.00	2.87	5.46	4.61
18	III	.3779	400	3.30	4.66	1.67	5.10	3.48
			∞	*	*	*	*	*
			200	4.27	21.13	1.99	21.12	4.35
4	IV	.2891	400	2.35	21.02	1.15	21.10	2.46
			∞	**		**		**
			200	4.45	20.32	1.89	20.51	4.51
9	IV	.2889	400	2.21	20.19	1.06	20.31	2.32
			∞	**		**		**
			200	4.33	15.86	1.65	16.10	4.35
14	IV	.2874	400	2.02	15.46	0.77	15.80	2.10
			∞	**		**		**
			200	4.32	14.28	1.78	14.50	4.41
19	IV	.2865	400	2.09	13.94	0.73	14.47	2.16
			∞	**		**		**
			200	3.41	26.35	1.78	26.44	3.44
5	V	.2328	400	2.10	26.67	0.94	26.81	2.25
			∞	**		**		**

*Significantly greater than chance at α=.001 for n=400.
** " " " " " α=.001 for n=200.

TABLE 5.37 (CONTINUED)

Pair	Group	Optimum	n	Full	First	Second	LDF	Matusita
			200	3.45	25.97	1.67	26.04	3.50
10	V	.2328	400	2.09	25.69	0.95	25.77	2.21
			∞	**		**		**
			200	3.17	20.88	1.09	21.51	3.21
15	V	.2327	400	1.56	20.62	0.66	21.08	1.59
			∞	**		**		**
			200	3.12	21.05	1.07	21.23	3.18
20	V	.2328	400	1.47	21.00	0.57	21.10	1.54
			∞	**		**		**

*Significantly greater than chance at α=.001 for n=400.
** " " " " " α=.001 for n=200.

TABLE 5.38

MEAN CORRELATION BETWEEN OBSERVED LOG LIKELIHOOD RATIOS AND TRUE LOG LIKELIHOOD RATIOS BASED ON 100 MONTE CARLO TRIALS FOR POPULATION PAIRS WITH $d_p \leq .1$

Pair	Group	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
		200	.0172	.6528	.2114	.4222	.0537	.001
6	I	400	.1203	.8003	.3408	.6005	.1487	-
		∞	1.0	.9790	1.0	.9790	1.0	
		200	.1353	.9629	.6009	.9281	.2650	.002
11	I	400	.2094	.9524	.5584	.8875	.2615	-
		∞	1.0	.9789	1.0	.9789	1.0	
		200	.5413	.0562	.7645	.0379	.5399	.095
7	II	400	.7501	.0619	.8148	.0518	.7793	.075
		∞	1.0	.0636	1.0	.0636	1.0	
		200	.6126	.2320	.7515	.1821	.6660	.093
12	II	400	.8040	.2366	.8075	.1870	.8091	.073
		∞	1.0	.2468	1.0	.2468	1.0	
		200	.6892	.2901	.7795	.2046	.7107	.095
17	II	400	.7298	.2901	.8031	.2316	.7696	.073
		∞	1.0	.3033	1.0	.3033	1.0	
		200	.2268	00	.5197	00	.2970	.012
3	III	400	.4407	00	.7585	00	.5010	.001
		∞	1.0	.0402	1.0	00	1.0	
		200	.3954	.0431	.7429	.0248	.4497	.012
8	III	900	.4259	.0327	.6658	.0191	.4604	.001
		∞	1.0	.0666	1.0	.0666	1.0	
		200	.3888	.2405	.5713	.1874	.4842	.010
13	III	400	.5512	.2616	.7771	.2408	.6387	.001
		∞	1.0	.2789	1.0	.2789	1.0	
		200	.2056	.3383	.7384	.3066	.3730	.011
18	III	400	.5517	.3379	.6945	.3113	.6488	-
		∞	1.0	.3545	1.0	.3545	1.0	
		200	.5598	00	.7135	00	.6097	.093
4	IV	400	.7252	00	.8473	00	.7583	.073
		∞	1.0	-.0064	1.0	-.0064	. 1.0	
		200	.6200	-.0012	.6928	-.0007	.6513	.094
9	IV	400	.7446	.0002	.8037	.0001	.7897	.071
		∞	1.0	-.0018	1.0	-.0018	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

TABLE 5.38 (CONTINUED)

Pair	Group	n	Full	First	Second	LDF	Matusita	Proportion neg.est.[a]
		200	.6040	-.0128	.6513	-.0103	.6991	.074
14	IV	400	.6217	-.0126	.7310	-.0102	.6734	.035
		∞	1.0	-.0139	1.0	-.0139	1.0	
		200	.4886	-.0209	.6107	-.0179	.5895	.071
19	IV	400	.5942	-.0209	.6837	-.0173	.6589	.030
		∞	1.0	-.0227	1.0	-.0227	1.0	
		200	.7850	-.0017	.7664	-.0008	.7817	.167
10	V	400	.9249	-.0054	.8890	-.0041	.9134	.144
		∞	1.0	-.0600	1.0	-.0600	1.0	
		200	.7479	-.0405	.8165	-.0267	.7949	.134
15	V	400	.7135	-.0469	.7374	-.0431	.7543	.084
		∞	1.0	-.0476	1.0	-.0476	1.0	
		200	.7457	-.0695	.7065	-.0537	.7740	.128
20	V	400	.6381	-.0743	.7612	-.0704	.6975	.071
		∞	1.0	-.0754	1.0	-.0754	1.0	

[a]The average proportion of times $\pi_i(x;[2])$ was less than zero.

accomplish this, the criterion proposed by Press (1972) was used which tests H_o: the number of correct classifications (hits) took place at random versus H_1: the discrimination procedure does better than chance. An asterisk (*) indicates that the classification procedure did better than chance for n=400, while a double asterisk (**) implies significance for n=200.

Note, there was no reason to initiate sampling experiments for population pair 1 since it is well known that under these conditions $\alpha = \alpha[*] = \bar{\Lambda}[*] = 1/2$. In fact, the equality holds for all cases in which the $\Pi = 1/2$, $p_{1j} = p_{2j}$, and $r_1(jk) = r_2(jk)$.

The tables reveal the linear models to be better than the other procedures in Group I population pairs (those with all $r_1(jk) = .10$). This is not surprising considering the results of the previous experiments wherein the performance of the first and LDF procedures was quite satisfactory for populations with moderate correlations. Also, the performance of the second procedure, for these population pairs, was somewhat worse than the first or LDF but better than the full and Matusita procedures. Note, however, for Group I population pairs none of the discrimination procedures were able to classify individuals better than chance assignment. With small mean differences and similar correlation structures, the underlying distributions have considerable overlap and, hence, it is not surprising that the use of each procedure resulted in considerable misclassification error.

In Group II population pairs (those with $r_1(jk)=.10$ and $r_2(jk)=.30$) the performance of the linear models was significantly worse than the others. The use of the linear procedures resulted in theoretical errors and mean actual errors which were significantly greater than the optimum error and, in addition, the mean correlations for these procedures were significantly lower. The performance of the second procedure was, as in Group I population pairs, somewhat better than that of the full or Matusita. For both sample sizes, the second procedure resulted in mean actual non-error rates which were significantly better than chance assignment.

Similar conclusions were derived for Group III population pairs; however, the relatively poor performance of the linear models appeared less severe than for those population pairs considered in Group II. All of the classification procedures did better than chance assignment for population pair 18 at n=400. Note this population pair had the largest admissible mean difference, i.e., $d_p=.1$.

Group IV and Group V contained population pairs for which the correlation structures in the two populations were most different. Thus, for these examples it was possible to determine the ability of each procedure to utilize information of this kind. For all of the population pairs considered in these two groups, the full, second and Matusita procedures performed significantly better than either of the linear models. The use of the first or LDF procedures yielded significantly greater theoretical errors and mean actual errors than the other procedures. In addition, an inspection of the mean correlations reveal that both $\rho[1]$ and $\rho[\ell]$ were less than or equal to zero for all of these population pairs. Uniformly, the discriminatory power of the linear models was not better than chance assignment.

On the other hand, the full, second and Matusita procedures seem to be able to effectively incorporate information provided by disparate correlation structures. The tables reveal that the use of these procedures yielded actual non-error rates which were better than chance assignment. Further, consider population pair 5 containing identical mean structures. Here, neither the first nor the LDF could effectively discriminate and, in fact, the use of either of these procedures resulted in theoretical errors and actual errors greater than 50 percent. Note, however, the corresponding error rates for the full, second or Matusita procedures were approximately one-half the magnitude.

To further illustrate the severe limitations the linear models when d_p is small, Figure 5.16 plots the optimum error and the theoretical errors for these procedures with $p_{1j}=p_{2j}=.5$. For this example, all $r_1(jk)$ were set to .10, while the correlation terms in population 2

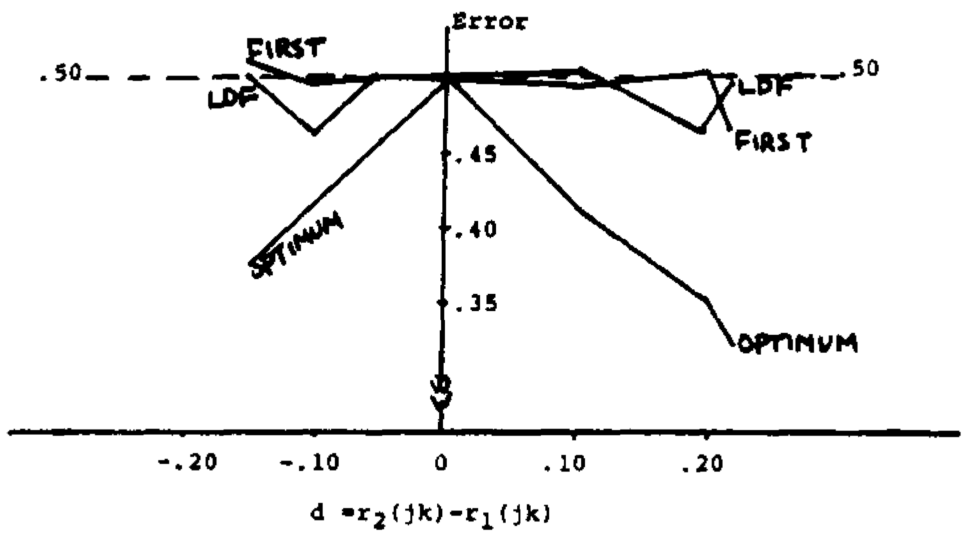

Figure 5.16 - Optimum and Theoretical Errors for $p_{1j}=p_{2j}=.50$; $j=1,2,\ldots\ldots,6$ and $r_1(jk)\neq r_2(jk)$.

were varied in a stepwise manner such that $r_2(jk)=$ $-.05$, 0, .05, .10, .20 and .30.

The figure reveals that neither the first nor the LDF procedures could effectively discriminate even when the correlation structures in the two populations were greatly different. The performance of the linear models was quite poor as evidenced by the especially high theoretical errors.

It would be interesting to determine whether similar results hold for small d_p with the added restriction of identical variance-covariance matrices. To accomplish this, two additional experiments were conducted with $d_p=.04$ and .1 wherein the correlation structures were varied such that the covariances in the two populations were the same. In the first experiment, $d_p=.04$ with $p_{2j}=1-p_{1j}$ and therefore all $r_1(jk)$ were set equal to all $r_2(jk)$. Here it is expected that none of the discrimination procedures will be able to effectively discriminate because of the considerable overlap in the population distributions. In the second experiment, however, $r_1(jk)\neq r_2(jk)$ since $p_{2j}=1-p_{1j}$ and values were assigned to the correlation terms in such a way as to maintain $S_1=S_2$.

Summary results for the mean increase in actual error at n=400 are given in Tables 5.39 and 5.40. The results for the cases with $d_p=.1$ appear quite similar to those obtained in previous sampling. The first and LDF procedures again performed relatively poorly except in

TABLE 5.39

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT)
BASED ON 100 MONTE CARLO TRIALS FOR
d_p=.04 WITH p_{1j}=.48,p_{2j}=.52, j=1,2,....,6 and n=400

r	Optimum Error	Full	First	Second	LDF	Matusita
0	.4625	2.48	0.68	1.72	0.71	2.50
.1	.4700	1.82	0.42	1.30	0.78	1.81
.2	.4701	1.67	0.85	1.16	0.99	1.67
.3	.4551	2.37	2.75	1.69	2.55	2.36
.33	.4507	2.62	3.22	1.85	2.92	2.60

TABLE 5.40

MEAN INCREASE IN ACTUAL OVER OPTIMUM ERROR (IN PERCENT)
BASED ON 100 MONTE CARLO TRIALS FOR
d_p=.1 WITH p_{1j}=.4, p_{2j}=.5, j=1,3.....,6 AND n=400

Correlations r_1(jk)	r_2(jk)	Optimum Error	Full	First	Second	LDF	Matusita
0	0	.3997	3.47	0.65	1.68	0.70	3.50
.1042	.1000	.4232	2.58	0.24	1.43	0.50	2.65
.2083	.2000	.4236	2.06	2.16	1.13	1.79	2.11
.3125	.3000	.3843	3.18	8.12	1.83	6.45	3.15
.3438	.3333	.3715	3.11	10.07	1.91	7.90	3.05

populations containing moderate correlations. For cases with d_p=.04 results were as expected. Uniformly, all procedures showed a marked inability to discriminate and although the mean increases in actual error for the first and LDF procedures were greater than those of the other procedures at large values of r, their relative performance was not significantly different from the others.

<u>5.31 - Conclusions - Question 3</u>: The results of the Monte Carlo sampling showed that:

1. The performance of the full, second and Matusita procedures was significantly better than either the first or LDF for all population pairs except those contained in Group I. However, for Group I population pairs none of the classification procedures yielded mean actual non-error rates which were significantly different from chance assignment.

2. The linear models should not be used whenever mean differences are slight. If $d_p \leq .1$, then the discriminatory power of the linear models was not better than chance assignment and it is likely that the use of these procedures on populations containing large correlations will yield theoretical errors and mean actual errors which are significantly greater than the optimum error.

3. It was rather apparent that the full, second and Matusita procedures can better utilize the information provided by disparate correlation structures. The ability

of these procedures to discriminate on the basis of correlation structure was clearly demonstrated for population pairs contained in Groups IV and V. In these groups the full, second and Matusita procedures did better than chance assignment across all population pairs.

5.4 - QUESTION 4 RESULTS

> What is the behavior of the Martin and Bradley model(s) under a variety of population structures? How does the performance of this model compare to the other discrimination procedures?

The probability model(s) suggested by Martin and Bradley has been developed in Section 4.2. Recall, the model depends on a set S of orthogonal polynomials $\phi_\gamma(x)$ on Ω with $h_s(a^{(i)},x)=\sum_{\gamma\epsilon\Gamma_s} a_j^{(i)}\phi_\gamma(x)$ being a distinguishing vector of parameters for θ_i i=1,2. The polynomials $\phi_\gamma(x)$, take on the values +1 and -1 and are analogous to the independent variables of a 2^m factorial design in the analysis of variance model. Hence, the model for the conditional probability mass function in the form $\pi_i(x)=f(x)[1+h(a^{(i)},x)]$, $i=1,2....,m$, $f(x)\geq 0$, $x\epsilon\Omega$ offers the unique advantage of isolating the joint contribution of variables to discrimination. Note the above model is based on the complete set of 2^m polynomials $\phi_\gamma(x)$, $\gamma\epsilon\Gamma_m$; however, several variations on this theme are possible. For example, the model can be approximated by including only lower order terms such as main effects and first order interactions or a lower set of polynomials S, where $S<m$,

can be utilized to estimate the $\pi_i(x)$. Obviously the principal advantage of such approximations lie in their ability to reduce the number of parameters to be estimated.

Although potentially useful, the literature at present lacks any investigation of the performance of these models. Therefore, the sampling experiments described below were specifically designed to examine the behavior of these models under a variety of population structures. Rather than merely focusing attention on the performance of the complete model, the sampling considers two additional specifications. The following briefly describes the models:

1. <u>Complete Model</u>: $\pi_i^C(x)=f(x)(1+h_C(a^{(i)},x))$; this model is based upon all polynomial terms up to and including order m.

2. <u>Incomplete Model</u>: $\pi_i^I(x)=f(x)(1+h_I(a^{(i)},x))$; in this model only main effects and first order interactions are included in the expansion of $h(a^{(i)},x)$.

3. <u>Reduced Model</u>: $\pi_i^R(x)=f(x)(1+h_R(a^{(i)},x))$; this model results from using a set of lower polynomials S in the estimation of $\pi_i(x)$. In all cases, a second order model is assumed, i.e., S=2.

Monte Carlo sampling experiments were accomplished for d_p=.1 and .4 where

$p_{1j}=(.4,.4,.4,.4,.4,.4)$, $p_{2j}=(.5,.5,.5,.5,.5,.5)$;
and

$p_{1j}=(.2,.2,.2,.2,.2,.2)$, $p_{2j}=(.6,.6,.6,.6,.6,.6)$
under two different sets of correlation patterns. The first set considered population pairs with $r_1(jk)\neq r_2(jk)$ for all $j\neq k$, $j,k=1,2,\ldots,6$. Here all correlation terms in population 1 were set to .10, while the correlation terms in population 2 were varied in a stepwise manner such that $r_2(jk)=-.05,0,.05,.10,.20,.30$ and $.33$. Stated differently, $d_r=r_2(jk)-r_1(jk)$ assumed values between -.15 and .23. The second set of correlation patterns was such that $r_1(jk)=r_2(jk)=r$. For these population pairs, r assumed the values 0,.10,.20,.30 and .33.

A measure of performance for the orthogonal polynomial models had to be developed in terms of the mean apparent error since the theoretical errors were intrackable. In general, the theoretical error can be found by specifying a_γ in lieu of $\{p_j\}$ and $\{q_j\}$; however this modification could not be easily handled with the existing sampling methodology and therefore only mean apparent errors are available. In addition to the complete, incomplete and reduced models, the section also presents results for the full and sccond procedures so as to provide a point of comparison.

Summary results for the two sets of correlation patterns with $d_p=.1$ and .4 are presented in Tables 5.41 and 5.42. The first point to note from the tables is the

TABLE 5.41

OPTIMUM ERROR AND MEAN APPARENT ERRORS BASED ON 100 MONTE CARLO TRIALS FOR

$r_1(jk) \neq r_2(jk)$ WITH d_p=.1 AND .4; [p_{1j}=.4, p_{2j}=.5] AND

[p_{1j}=.2, p_{2j}=.6], j=1,2,......6

d_r			Mean Apparent Errors			Incomplete	
$[r_2(jk)-r_1(jk)]$	n	Optimum Error	Full	Second	Complete	Main Effects	First Order Interaction
d_p=.1							
	200		.2942	.3392	.2942	.4241	.3410
-.15	400	.3752	.3269	.3562	.3269	.4318	.3592
	200		.3091	.3610	.3090	.4114	.3556
-.10	400	.3992	.3465	.3776	.3466	.4217	.3792
	200		.3202	.3712	.3202	.4040	.3678
-.05	400	.4114	.3537	.3878	.3537	.4153	.3884
	200		.3231	.3808	.3230	.4057	.3751
0	400	.4232	.3579	.3960	.3579	.4139	.3959
	200		.3222	.3725	.3221	.3953	.3632
.10	400	.4054	.3548	.3877	.3548	.4012	.3840
	200		.2984	.3405	.2993	.3667	.3145
.20	400	.3736	.3284	.3522	.3270	.3621	.3186
	200		.2902	.3272	.2890	.3457	.2891
.23	400	.3575	.3162	.3398	.3131	.3375	.2910

TABLE 5.41 (CONTINUED)

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Mean Apparent Errors: Full	Mean Apparent Errors: Second	Mean Apparent Errors: Complete	Incomplete: Main Effects	Incomplete: First Order Interaction
$d_p=.4$							
	200		.1276	.1458	.1268	.2026	.1774
-.15	400	.1527	.1413	.1498	.1404	.2003	.1801
	200		.1421	.1615	.1408	.2070	.1849
-.10	400	.1698	.1562	.1670	.1553	.2080	.1870
	200		.1506	.1738	.1518	.2169	.1930
-.05	400	.1870	.1663	.1792	.1686	.2159	.1960
	200		.1615	.1784	.1633	.2214	.1991
0	400	.2043	.1783	.1844	.1786	.2224	.2027
	200		.1771	.2093	.1804	.2439	.2331
.10	400	.2389	.1986	.2209	.1972	.2408	.2344
	200		.1863	.2225	.1865	.2738	.2905
.20	400	.2402	.2044	.2293	.2066	.2646	.2978
	200		.1868	.2204	.1865	.2876	.3069
.23	400	.2367	.2038	.2273	.2056	.2854	.3180

TABLE 5.42

OPTIMUM ERROR AND MEAN APPARENT ERRORS BASED ON 100 MONTE CARLO TRIALS

FOR $r_1(jk)=r_2(jk)=r$ WITH $d_p=.1$ AND $.4$; $[p_{1j}=.4,\ p_{2j}=.5]$ AND

$[p_{1j}=.2, p_j=.6]$, $j=1,2,\ldots\ldots6$

r	n	Optimum Error	Mean Apparent Errors: Full	Mean Apparent Errors: Second	Mean Apparent Errors: Complete	Incomplete: Main Effects	Incomplete: First Order Interaction
$d_p=.1$							
0	200		.3192	.3660	.3196	.3923	.3631
	400	.3997	.3534	.3824	.3521	.3953	.3811
.10	200		.3231	.3808	.3239	.4019	.3720
	400	.4232	.3579	.3960	.3618	.4141	.3980
.20	200		.3277	.3824	.3297	.4246	.3844
	400	.4240	.3602	.3958	.3602	.4356	.4135
.30	200		.3251	.3664	.3266	.4365	.3797
	400	.3863	.3551	.3815	.3444	.4464	.3907
.33	200		.3245	.3618	.3276	.4348	.3723
	400	.3745	.3478	.3698	.3467	.4354	.3702

TABLE 5.42 (CONTINUED)

r	n	Optimum Error	Mean Apparent Errors: Full	Second	Complete	Incomplete: Main Effects	First Order Interaction
$d_p = .4$							
	200		.1207	.1340	.1216	.2018	.1811
0	400	.1390	.1323	.1379	.1318	.2001	.1813
	200		.1615	.1784	.1605	.2223	.1986
.10	400	.2043	.1783	.1844	.1803	.2238	.2033
	200		.1780	.2177	.1782	.2890	.2933
.20	400	.2363	.2007	.2253	.2000	.2913	.3033
	200		.1384	.1746	.1401	.4247	.2507
.30	400	.1786	.1565	.1763	.1571	.4341	.2614
	200		.1162	.1301	.1159	.4342	.2116
.33	400	.1349	.1255	.1330	.1259	.4446	.2141

absence of results for the reduced model. The sampling experiments determined that the performance of this procedure was extremely optimistically biased for all parameter values whenever the original values for the coefficients were utilized. The apparent errors obtained by using the original values for a_γ were less than 10 percent across all of the population structures. This result supports the Martin and Bradley contention that an iterative procedure should be used for estimating the coefficients a_γ when employing a reduced model. Also, note the incomplete models report two apparent errors. The first apparent error, "main effects", was derived from including only main effects in the expansion of $(h(a^{(i)}, x)$, while the apparent error corresponding to "first order interaction" includes both main effects plus first order interactions. Both approximations are reported so as to determine whether lower order models which involve fewer parameter estimates can yield satisfactory results.

As expected, the use of the full, second and complete procedures results in mean apparent errors which are less than the optimum error. This observation merely illustrates the fact that the optimum error provides an upper bound for the mean apparent error. This does not, however, necessarily hold for the incomplete models since often the use of an approximation may destroy the orthogonality property. Hence, it is more likely that the mean apparent errors for these models will vary greatly across the different population structures. It is apparent from the tables that the full multinomial and complete models give near identical results. This is not too surprising since the complete model requires the estimation of approximately the same number of parameters as does the full.

The relative performance of each procedure can be evaluated by utilizing the mean apparent error bias which is simply computed by taking the difference between optimum error and mean apparent error. Tables 5.43 and 5.44 present summary results for each of the population structures.

The tables reveal that the performance of the full, second and complete models was affected by larger sample sizes. Uniformly, across each population structure, the mean apparent error bias decreased as n was increased. It is well known that the expected apparent error approaches the optimum error as n becomes large and hence this result is not surprising. The behavior of the mean apparent error for the incomplete models was far less precise since with these models the mean apparent error bias was capable of reversing both direction and sign as d_r and n were varied. Obviously, this does present some difficulties when attempting to talk to general trends and conclusions across different population structures. To eliminate this possible source of contradiction remarks will be restricted to cases where n=400 unless otherwise stated. To better judge the

TABLE 5.43

MEAN APPARENT ERROR BIAS (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS FOR $r_1(jk) \neq r_2(jk)$, WITH d_p=.1 AND .4; [p_{1j}=.4, p_{2j}=.5) AND [p_{1j}=.2, p_{2j}=.6], j=1,2,......6

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Mean Apparent Error Bias: Full	Second	Complete	Incomplete: Main Effects	First Order Interaction
d_p=.1							
	200		8.10	3.60	8.10	-4.89	3.42
-.15	400	.3752	4.83	1.90	4.83	-5.66	1.60
	200		9.01	3.82	9.01	-1.22	4.36
-.10	400	.3992	5.27	2.16	5.27	-2.25	2.00
	200		9.12	4.02	9.12	.74	4.36
-.05	400	.4114	5.77	2.36	5.77	-.39	2.30
	200		10.01	4.24	10.02	1.75	4.81
0	400	.4232	6.53	2.72	6.53	.99	2.73
	200		8.32	3.29	8.33	1.01	4.22
.10	400	.4054	5.06	1.77	5.06	.42	2.14
	200		7.52	3.31	7.43	.69	5.91
.20	400	.3736	4.52	2.14	4.66	1.15	5.50
	200		6.73	3.03	6.85	1.18	6.84
.23	400	.3575	4.13	1.77	4.44	2.00	6.65

TABLE 5.43 (CONTINUED)

d_r $[r_2(jk)-r_1(jk)]$	n	Optimum Error	Mean Apparent Error Bias Full	Mean Apparent Error Bias Second	Mean Apparent Error Bias Complete	Incomplete Main Effects	Incomplete First Order Interaction
$d_p=.4$							
	200		2.51	.69	2.59	-4.99	-2.47
-.15	400	.1527	1.14	.29	1.23	-4.76	-2.74
	200		2.77	.83	2.90	-3.72	-1.51
-.10	400	.1698	1.36	.28	1.45	-3.82	-1.92
	200		3.64	1.32	3.52	-2.99	-.60
-.05	400	.1870	2.07	.78	1.84	-2.89	-.90
	200		4.28	2.59	4.10	-1.71	.52
0	400	.2043	2.60	1.99	2.57	-1.81	.16
	200		6.18	2.96	5.85	-.50	.58
.10	400	.2389	4.03	1.80	4.17	-.19	.45
	200		5.39	1.77	5.37	-3.36	-5.03
.20	400	.2402	3.58	1.09	3.36	-2.24	-5.76
	200		4.99	1.63	5.02	-5.09	-7.02
.23	400	.2367	3.29	.94	3.11	-4.87	-8.13

TABLE 5.44

MEAN APPARENT ERROR BIAS (IN PERCENT) BASED ON 100 MONTE CARLO TRIALS

FOR $r_1(jk)=r_2(jk)=r$, WITH $d_p=.1$ AND .4; $[p_{1j}=.4,\ p_{2j}=.5]$ AND

$[p_{1j}=.2, p_{2j}=.6]$, $j=1,2,\ldots\ldots 6$

r	n	Optimum Error	Mean Apparent Error Bias: Full	Mean Apparent Error Bias: Second	Mean Apparent Error Bias: Complete	Incomplete: Main Effects	Incomplete: First Order Interaction
$d_p=.1$							
	200		8.05	3.37	8.01	.74	3.66
0	400	.3997	4.63	1.73	4.76	.44	1.86
	200		10.01	4.24	9.93	2.13	5.12
.1	400	.4232	6.53	2.72	6.14	.91	2.52
	200		9.63	4.16	9.43	-.06	3.96
.2	400	.4240	6.38	2.82	6.38	-1.16	1.05
	200		6.12	1.99	5.97	5.02	.66
.3	400	.3863	3.12	.48	3.69	6.01	.44
	200		5.00	1.27	4.69	-6.03	.22
.33	400	.3745	2.67	.47	2.78	-6.09	.43

TABLE 5.44 (CONTINUED)

r	n	Optimum Error	Mean Apparent Error Bias: Full	Second	Complete	Incomplete: Main Effects	First Order Interaction
d_p=.4							
0	200		1.83	.50	.74	-6.28	-4.21
	400	.1390	.67	.11	.72	-6.11	-4.23
.1	200		4.28	2.59	9.38	-1.80	.57
	400	.2013	2.60	1.99	2.40	-1.95	.10
.2	200		5.83	1.86	5.81	-5.27	-5.70
	400	.2363	3.56	1.10	3.63	-5.50	-6.70
.3	200		4.02	.40	4.02	-24.61	-7.21
	400	.1786	2.21	.23	2.21	-25.55	-8.28
.33	200		1.87	.48	1.90	-29.93	-7.67
	400	.1349	.94	.19	.90	-30.97	-7.92

superiority of one procedure over another, the mean apparent errors for each procedure were ranked according to their magnitude; summary results are given in Tables 5.45 and 5.46.

The following observations were derived from Tables 5.43 and 5.45 which present the mean apparent error bias for each procedure and its corresponding rank for population structures characterized by $r_1(jk) \neq r_2(jk)$.

1. The magnitude and direction of the mean apparent error bias for the full, second and complete models were fairly consistent over d_r for both d_p=.1 and .4 It is also apparent that smaller biases are found in those population structures with relative smaller optimum errors. That is, the mean apparent error bias is a decreasing function of the optimum error. This result is not unusual, for it follows from the known behavior of the expected apparent error.

2. The behavior of the incomplete models was, on the other hand, quite different from either the full, second or complete models. Rather than decreasing as $|d_r|$ increased, the magnitude of the bias for both the main effects and first order interaction models increased with larger $|d_r|$. Also, there were several population pairs for which the bias with d_p=.4 was shown to be greater than that with d_p=.1.

TABLE 5.45

RANKS OF THE MEAN APPARENT ERROR BIAS FOR $r_1(jk) \neq r_2(jk)$

WITH d_p=.1 AND .4; [p_{1j}=.4, p_{2j}=.3] AND

[p_{1j}=.2, p_{2j}=.6], j=1,2,.....6

d_r [$r_2(jk)-r_1(jk)$]	n	Rankings: Full	Rankings: Second	Rankings: Complete	Rankings: Incomplete: Main Effects	Rankings: Incomplete: First Order Interaction
d_p=.1						
-.15	200	4.5	2	4.5	3	1
	400	3.5	2	3.5	5	1
-.10	200	4.5	2	4.5	1	3
	400	4.5	2	4.5	3	1
-.05	200	4.5	2	4.5	1	3
	400	4.5	3	4.5	1	2
0	200	4	2	5	1	3
	400	4.5	2	4.5	1	3
.10	200	4	2	5	1	3
	400	4.5	2	4.5	1	3
.20	200	5	2	4	1	3
	400	3	2	4	1	5
.23	200	3	2	5	1	4
	400	3	2	4	1	5

TABLE 5.45 (continued)

d_r $[r_2(jk)-r_1(jk)]$	n	Full	Second	Complete	Incomplete Main Effects	Incomplete First Order Interaction
$d_p=.4$						
-.15	200	3	1	4	5	2
	400	2	1	3	5	4
-.10	200	3	1	4	5	2
	400	2	1	3	5	9
-.05	200	5	2	4	3	1
	400	4	2	3	5	1
0	200	5	3	4	2	1
	400	5	3	4	2	1
.10	200	5	3	4	1	2
	400	4	3	5	1	2
.20	200	5	1	4	2	3
	400	4	1	3	2	5
.23	200	2	1	3	4	5
	300	3	1	2	4	5

TABLE 5.46

RANKS OF THE MEAN APPARENT ERROR BIAS FOR $r_1(jk)=r_2(jk)=r$

WITH d_p=.1 AND .4; $[p_{1j}=.4, p_{2j}=.5]$ AND

$[p_{1j}=.2,\ p_{2j}=.6]$, j=1,2.....6

Rankings

r	n	Full	Second	Complete	Incomplete: Main Effects	Incomplete: First Order Interaction
d_p=.1						
0	200	5	2	4	1	3
	400	4	2	5	1	3
.1	200	5	2	4	1	3
	400	5	3	4	1	2
.2	200	5	3	4	1	2
	400	4.5	3	4.5	2	1
.3	200	5	1	4	3	2
	400	3	2	4	5	1
.33	200	4	2	3	5	1
	400	3	2	4	5	1

TABLE 5.46 (CONTINUED)

r	n	Full	Second	Complete	Incomplete: Main Effects	Incomplete: First Order Interaction
d_p=.4						
	200	3	1	2	5	4
0	400	2	1	3	5	4
	200	4	3	5	2	1
.1	400	5	3	4	2	1
	200	5	1	4	2	3
.2	400	2	1	3	4	5
	200	2.5	1	2.5	5	4
.3	400	2.5	1	2.5	5	1
	200	2	1	3	5	4
.33	400	3	1	2	5	4

3. For moderate values of d_r, it was expected that the incomplete models would do well since they involve fewer parameter estimates and should be able to satisfactorily characterize the underlying population distributions. Although the tables reveal this to be true, it was rather surprising to note the relatively "good" performance of the main effects model across all d_r values with d_p=.1. For these population pairs, the main effects model had the lowest mean apparent error bias in five of the seven population structures. However, in population structures with d_p=.4, the main effects model did progressively worse as $|d_r|$ increased.

Tables 5.44 and 5.46 present summary results for population structures with $r_1(jk)=r_2(jk)=r$. A study of the tables yields the following conclusions:

1. The behavior of the incomplete models was again quite erratic across population structures with different main vectors. For example, with d_p=.1 and r=0, the main effects model yielded the smallest mean apparent error bias, while with d_p=.4 and r=0, it yielded the largest bias. A similar result was found for the first order interaction model. Here, with d_p=.1 and r=.33 use of this model yielded relatively small bias, while relatively large bias was obtained with d_p=.4 and r=.33.

2. The use of the main effects model with large r yields relatively large negative biases. This was especially true for those population pairs with d_p=.4. Note these values of r correspond to structures for which reversals are likely to occur.

3. The use of the complete model again gave near identical results as the full multinomial. Also, smaller mean apparent error biases for the full, second and complete models were found in those population structures having relatively smaller optimum error.

The behavior of the incomplete models proved rather difficult to describe. There is a natural tendency to equate means with main effects and similarly, correlations with first order interactions. However, the Martin and Bradley parameters when using a reduced form of the complete model are not identical to the Bahadur reparametization used to generate the Monte Carlo samples. Hence, to interpret the incomplete models in terms of means and correlations instead of main effects and interactions is understandably difficult.

5.14.1 - Conclusions: Question 4. The following conclusions are based on the results of the Monte Carlo sampling experiments already described.

1. A reduced model should not be used unless an iterative procedure is employed which modifies the original coefficients a_γ. Use of a reduced model without an iterative

solution will result in apparent errors which are severely optimistically biased.

2. The incomplete models were shown to result in satisfactory mean apparent errors under several different kinds of population structures. Although there was some difficulty in interpreting these models with respect to means and correlations, it is nevertheless safe to conclude that they will do well on population structures containing moderate correlations.

3. The evidence suggests that the behavior of the complete model parrots that of the full multinomial quite closely. Also, it appears that the full or complete models should be used on structures containing larger correlation terms.

5.5 - QUESTION 5 RESULTS

> What are the effects of unequal sample size on the Matusita model? With equal sample sizes the classification rule for this model can be shown to be equivalent to the usual non-parametric rule; however, the effect of unequal sample sizes with equal priors is unclear.

The classification rule derived from the work of Kameo Matusita (1952, 1955, 1956) has been developed in Section 4.2. The rule considers classifying an observation Z=z into one of two distributions, F or G, on the basis of the empirical cumulative distribution function(s), S_{n+1}, (S_{m+1}) when z is assumed to be a sample point generated by F(G). To recapitulate, let the possible multinomial states be given as $E_1, E_2, \ldots, E_k$ then the rule classifying the observation Z=z, signifying state E_j, into F is given by

$$\sum_{\substack{i=1\\ i\neq j}}^{k}\left[\frac{n_i}{n}\cdot\frac{m_i}{m+1}\right]^{1/2} + \left[\frac{n_j}{n}\cdot\frac{m_j+1}{m+1}\right]^{1/2} \geq \sum_{\substack{i=1\\ i\neq j}}^{k}\left[\frac{n_i}{n+1}\cdot\frac{m_i}{m}\right]^{1/2} + \left[\frac{n_j+1}{n+1}\cdot\frac{m_j}{m}\right]^{1/2}$$

where n and m are respective sample sizes of the random samples taken from F and G. With equal sample sizes, n=m, the rule reduces to the usual non-parametric rule: Classify Z=z into F if

$$\frac{n_j}{n} \geq \frac{m_j}{m},$$

if the random variable Z signifies state E_j.

In all of the previous sampling experiments $n_1=n_2$ and therefore in many of the examples it was not necessary to report results for the Matusita model since under this condition the classification rule for this model can be cast as a sample-based likelihood ratio test which is equivalent to the full multinomial procedure. Nevertheless, the sampling experiments did show the mean correlations for this procedure to be somewhat better than the full for most values of p_{ij}, $d_r(r)$ and n. In part, the explanation for this result may be due to the way in which the Matusita model handles empty cells. Recall, this procedure is capable of utilizing the information available from other states which is not the case using the full procedure.

With unequal sample sizes, however, the Matusita classification rule does not reduce to a likelihood ratio test and hence results are not clear. Thus, further Monte Carlo sampling experiments were initiated on population structures for which $n_1 \neq n_2$. In all of the following experiments the sample size in population 1 was fixed at 200, while the sample size in population 2 was set at levels of 300 and 400, corresponding to one and half, and twice the number of observations in population 1. The sampling was accomplished under three sets of correlation patterns. The first set contained population pairs with only one correlation term different from zero ($r_2(13)$). For these population pairs all $r_1(jk)=0$, while $r_2(13)$ was incremented in a stepwise manner from -.6 to .6. The second set of correlation patterns were such that all $r_1(jk) \neq r_2(jk)$. Here $r_2(jk)$ assumed that values -.05, .10, .20, .30 and .33, while all $r_1(jk)$ were set to .10. The third set of correlation patterns restricted the correlation terms in population 1 to be equal to those in population 2, i.e., $r_1(jk)=r_2(jk)=r$ for all $j \neq k$. For population pairs considered in this set r assumed the values 0, .10, .20, .30 and .33. Initially, the values assigned to the marginal probability vectors in the two populations were taken to be

$$p_{1j}=(.4,.4,.4,.4,.4,.4) \quad , \; p_{2j}=(.5,.5,.5,.5,.5,.5).$$

However, in order to determine whether larger mean differences alter results, one additional sampling experiment was conducted with

$$p_{1j}=(.2,.2,.2,.2,.2,.2) \quad , \quad p_{2j}=(.6,.6,.6,.6,.6,.6),$$

in population structures characterized by $r_1(jk)=r_2(jk)=r$.

In all experiments performance was measured in terms of two criteria, the mean actual error and the mean apparent error. In order to judge performance, results were also tabulated for the full multinomial. Tables 5.47 through 5.50 give summary results. Note the tables present results for the case where $n_1=n_2=200$.

5.51 - Conclusions - Question 5: The following conclusions are based on the results of the series of Monte Carlo sampling experiments described above and displayed in Tables 5.47 through 5.50.

1. The performance of the Matusita model under the condition of unequal sample sizes appears to be quite satisfactory with respect to both the mean actual error and mean apparent error. Even when the sample size in one population was twice that of the other, the Matusita model and the full multinomial procedure gave near identical results across most values of $d_r(r)$ and d_p.

2. Although the classification rule associated with the Matusita model does not have the properties of a likelihood ratio test whenever $n_1 \neq n_2$, there is some evidence to suggest that this procedure may be particularly useful on population structures containing extreme correlation terms. Generally, the data are more sparse as the correlations in either population assume large positive or negative values.

TABLE 5.47

MEAN APPARENT AND MEAN ACTUAL ERRORS BASED ON 100 MONTE CARLO TRIALS WITH UNEQUAL SAMPLES FOR d_p=.1; p_{1j}=.4, p_{2j}=.5, j=1,2,....6 AND $r_1(13) \neq r_2(13)$

d_r [$r_2(13)$]	Mean Apparent Full	Mean Apparent Matusita	Mean Actual Full	Mean Actual Matusita
n_2=200				
-.6	.2841	.2841	.3803	.3803
-.5	.2895	.2895	.4022	.4022
-.4	.3015	.3015	.4190	.4191
-.3	.3098	.3098	.4314	.4315
-.2	.3131	.3131	.4401	.4401
-.1	.3183	.3183	.4452	.4452
0	.3149	.3.49	.4477	.4478
.1	.3162	.3162	.4475	.4474
.2	.3163	.3163	.4434	.4433
.3	.3120	.3120	.4346	.4347
.4	.3061	.3061	.4217	.4217
.5	.2964	.2964	.4045	.4044
.6	.2855	.2855	.3883	.3883
n_2=300				
-.6	.2874	.3738	.3803	.3727
-.5	.3030	.3943	.4022	.3931
-.4	.3112	.4121	.4190	.4117
-.3	.3188	.4255	.4314	.4254
-.2	.3243	.4343	.4401	.4352
-.1	.3256	.4421	.4452	.4433
0	.3266	.4443	.4477	.4466
.1	.3267	.4421	.4475	.4437
.2	.3286	.4359	.4434	.4376
.3	.3210	.4290	.4346	.4293
.4	.3132	.4167	.4217	.4166
.5	.3052	.3989	.4045	.3979
.6	.2926	.3827	.3883	.3823

TABLE 5.47 (continued)

d_r [$r_2(13)$]	Mean Apparent Full	Mean Apparent Matusita	Mean Actual Full	Mean Actual Matusita
n_2=400				
-.6	.2896	.2899	.3639	.3655
-.5	.3050	.3052	.3892	.3870
-.4	.3127	.3218	.4108	.4090
-.3	.3224	.3224	.4216	.4217
-.2	.3280	.3280	.4330	.4336
-.1	.3351	.3351	.4410	.4431
0	.3346	.3346	.4433	.4461
.1	.3343	.3343	.4240	.4441
.2	.3289	.3289	.4361	.4368
.3	.3252	.3252	.4251	.4255
.4	.3165	.3165	.4147	.4140
.5	.3103	.3144	.3982	.3963
.6	.2994	.2996	.3796	.3767

TABLE 5.48

MEAN APPARENT AND MEAN ACTUAL ERRORS BASED ON 100 MONTE CARLO TRIALS WITH UNEQUAL SAMPLES FOR d_p=.1, p_{1j}=.4, p_{2j}=.5, j=1,2,....6 AND $r_1(jk) \neq r_2(jk)$

d_r $[r_2(jk) \neq r_1(jk)]$	Mean Apparent Full	Mean Apparent Matusita	Mean Actual Full	Mean Actual Matusita
n_2=200				
-.15	.2942	.2942	.4169	.4169
-.10	.3092	.3091	.4397	.4397
-.05	.3202	.3202	.4519	.4519
0	.3231	.3231	.4577	.4578
.10	.3222	.3222	.4482	.4479
.20	.2984	.2984	.4058	.4047
.23	.2902	.2902	.3872	.3861
n_2=300				
-.15	.3029	.3029	.4121	.4100
-.10	.3193	.3193	.4354	.4366
-.05	.3292	.3292	.4479	.4495
0	.3333	.3333	.4546	.4560
.10	.3312	.3312	.4447	.4438
.20	.3072	.3072	.4027	.3995
.23	.2978	.2978	.3859	.3770
n_2=400				
-.15	.3070	.3075	.4091	.4000
-.10	.3257	.3257	.4334	.4346
-.05	.3376	.3376	.4472	.4486
0	.3389	.3389	.4530	.4540
.10	.3390	.3393	.4453	.4438
.20	.3092	.3105	.4030	.3990
.23	.3004	.3020	.3846	.3702

TABLE 5.49

MEAN APPARENT AND MEAN ACTUAL ERRORS BASED ON 100 MONTE CARLO TRIALS WITH UNEQUAL SAMPLES FOR d_p=.1; p_{1j}=.4, p_{2j}=.5, j=1,2....6 AND $r_1(jk)=r_2(jk)=r$

r	Mean Apparent Full	Mean Apparent Matusita	Mean Actual Full	Mean Actual Matusita
n_2=200				
0	.3192	.3192	.4483	.4484
.10	.3231	.3231	.4577	.4578
.20	.3277	.3277	.4539	.4539
.30	.3251	.3251	.4290	.4286
.33	.3478	.3478	.4200	.4190
n_2=300				
0	.3266	.3266	.4433	.4466
.10	.3338	.3338	.4546	.4554
.20	.3335	.3335	.4495	.4499
.30	.3329	.3330	.4270	.4243
.33	.3325	.3326	.4164	.4101
n_2=400				
0	.3346	.3346	.4433	.4461
.10	.3399	.3399	.4545	.4552
.20	.3408	.3408	.4491	.4494
.30	.3354	.3356	.4265	.4235
.33	.3371	.3375	.4151	.4095

TABLE 5.50

MEAN APPARENT AND MEAN ACTUAL ERRORS BASED ON 100 MONTE CARLO TRIALS WITH UNEQUAL SAMPLES FOR $d_p = .4$; $p_{1j} = .2$, $p_{2j} = .6$, $j=1,2,\ldots,6$ AND $r_1(jk) = r_2(jk) = r$

r	Mean Apparent Full	Mean Apparent Matusita	Mean Actual Full	Mean Actual Matusita
$n_2=200$				
0	.1207	.1207	.1720	.1720
.10	.1615	.1615	.2333	.2334
.20	.1780	.1780	.2683	.2684
.30	.1384	.1384	.2144	.2144
.33	.1162	.1162	.1529	.1521
$n_2=300$				
0	.1254	.1265	.1647	.1640
.10	.1669	.1677	.2268	.2258
.20	.1821	.1831	.2650	.2645
.30	.1437	.1445	.2904	.2052
.33	.1179	.1183	.1656	.1591
$n_2=400$				
0	.1266	.1294	.1632	.1618
.10	.1703	.1722	.2259	.2252
.20	.1870	.1891	.2630	.2625
.30	.1462	.1475	.2068	.2040
.33	.1198	.1209	.1638	.1604

Note, however, it was for these kinds of structures that the Matusita sample-based classification rule resulted in smaller mean actual error than the full multinomial and, hence, this procedure may offer special efficacy for sparse data.

5.6 - GENERAL CONCLUSIONS AND SUGGESTIONS

A study of the theoretical and optimum errors and examination of numerous Monte Carlo sampling experiments lead to the following general conclusions and suggestions:

1. The full, second and Matusita procedures should be used whenever the mean vectors in the two populations are similar and when it is suspected or known beforehand that the correlations are large. For the most part, the use of the second order procedure results in lower mean actual error than either the full or Matusita over a considerable range of values for the parameters $p_{ij}, r_1(jk)$ and n. However, it is advisable to note that sampling was initiated in the absence of third and higher order terms and, hence, the superiority of the second order procedure is likely to be somewhat overstated.

2. Neither the first nor LDF procedures should be used whenever mean differences are small or with highly correlated variables. Even for samples taken from populations having identical variance-covariance matrices, the use of the linear model results in significantly greater misclassification (on the average) than either the full, second, Matusita or complete models.

3. As expected, the performance of the first order procedure is quite similar to that of the LDF. These procedures should be used on population structures containing moderate correlation terms. For populations of this kind the use of either the first or LDF procedures should result in lower mean actual error than that of any other procedure.

4. The full multinomial and Matusita sample-based classification rule give near identical results for the unequal sample size with equal priors problems as well as for the case of equal sample sizes and equal priors. Based upon the preliminary results, the Matusita model seems potentially useful in cases where the data are sparse and with smaller sample sizes. For these kinds of population structures, the mean correlation between the observed log likelihood ratios and the time log likelihood ratios should be higher for the Matusita model than the full multinomial procedure.

5. The Martin and Bradley reduced model should not be used unless an iterative procedure is employed which modifies the original coefficients. The use of this model without an iterative solution will result in apparent errors which are severely optimistically biased.

6. The performance of the Martin and Bradley complete model is nearly identical to the full and Matusita procedures and should be used on populations having little sparseness

and high correlations. On the other hand, the incomplete models appear most appropriate when the correlation terms are moderate.

The next chapter considers an example with application to data on communication buyer behavior.

CHAPTER VI

APPLICATION TO DATA: RESEARCH FINDINGS AND IMPLICATIONS

This chapter considers an application of the various discrimination procedures to a communication buyer behavior data base. Demography was utilized as the focal point for describing and segmenting the markets. This was thought to be of extreme interest since many researchers are skeptical of demographic factors as determinants or even correlates of consumption behavior (e.g., Yankelowich, 1964; Frank, Massy and Boyd, 1967; Frank, 1968; Wells and Tigert, 1971).

Much of the criticisms levied against demography stem from empirical studies, especially on frequently purchased grocery products, which reveal poor performance on the part of demographic factors in explaining differences in brand loyalty, deal proneness or consumption patterns (Frank, Massy and Boyd, 1967; Frank, 1968; Frank, Massy and Lodhal, 1968). Almost uniformly, these studies have yielded low R^2 (coefficient of multiple determination) between demographic factors and any aspect of consumption behavior.

The occurrence of low R^2 in the majority of empirical studies cannot be disputed; however, the conclusion that demography is inadequate as a segmenting variable seems too strong a statement. There are several reasons which explain the relatively poor showing of demography. A number of these reasons fall under the broad umbrella of measurement problems. Generally, demographic data are at best discrete and often these variables are dichotomized, producing dummy variables. Clearly, to use statistical procedures which impose restrictions of normality, linearity and additivity on the relationship among variables must be viewed as being highly suspect. Also, given the fact that both the criterion variable and the independent variables are discrete, then it can be shown that the maximum value of R^2 may be considerably less than unity (Martin, 1973). Thus, it would seem that demography warrants further examination before proceeding to discard it as a determinant of consumption behavior.

Toward this end, this section seeks to determine whether the employment of alternative discrimination procedures--those which are more compatible with categorical data--offer additional insights into the utilization of demography in differentiating between groups of product users. Among the other major issues to be explored are:

1. How do the results of the Monte Carlo sampling experiments test on this data? Do error rates vary greatly among the discrimination procedures?

2. How should variables be coded? What are the effects of dichotomizing those questions with multiple response categories?

3. How can good subsets of variables be selected for the two-group multinomial classification problem? How well does the optimal subset of variables discriminate?

6.1 - FIELD STUDY DESIGN[1]

The data utilized in this chapter was provided by a major United States corporation. The corporation, which has a number of associated companies servicing the majority of product users, selected three metropolitan areas considered representative of their total user population. Taking into consideration the nature of the project and the associated costs of data collection, it was decided that a data base consisting of 450 respondents would be sufficient.

The three areas selected for sampling were Atlanta, Indianapolis and Los Angeles. The individuals selected from these areas resided in both the central city and surrounding suburban communities. In addition, an evenly dispersed quota sample based upon respondents' estimates

[1]This researcher did not participate in either the design or collection of the questionnaire; a data deck and documentation were made available.

of their product usage rates was chosen as the method of selection. Toward that end, there was selected a nearly equal number of respondents from each of the predetermined usage categories roughly corresponding to what would be called "heavy", "medium", and "light" users. The unique nature of the product category under consideration made it possible for the company to cross-validate each individual's usage classification in terms of actual units consumed.

Prior to the distribution of the questionnaire, an independent data collection firm completed a number of pretests of the questionnaire. Based upon this experience, a number of adjustments were made to the original questionnaire. The independent collection firm was also responsible for the distribution and pick up of all questionnaires.

In distributing the questionnaire, interviewers were instructed to select a number of central city and suburban communities within each of the sampling areas comprising a mix of lower, middle and upper socio-economic households. At the start of each call, interviewers screened potential respondents for the following:

1. that the household was a user of a specific communication product, and
2. that a member of the household would agree to complete the questionnaire and have the interviewer return to pick it up.

In order to ensure a minimum of 450 respondents each area was slightly over-sampled. A final usable data base of 464 respondents was secured. This constituted a response rate of over 90 percent of those who agreed to complete the questionnaire.

6.2 - DEFINITION AND MEASUREMENT OF VARIABLES

Initially, each household was instructed to rate their usage of the product under consideration in comparison to that of their friends and neighbors. In this way, each household was placed into either a heavy, medium, or light product user category. However, for the purposes of this analysis, it was decided to separate the 464 respondents into two populations, corresponding to heavy or light usage, on the basis of actual dollar usage rates. Based upon company objectives and prior research findings, the decision was to classify a household into the heavy group, population 1, if their usage rate per period was $5.00 or more, and into the light group, population 2, if usage rate for period was less than $5.00.

Although this study is restricted to examining the use of demographic data as a basis for describing market segments, respondents were asked to provide information relating to three broad areas:

1. activities, interests, and opinions (A.I.O);
2. product usage behavior; and
3. demographic data

It does not seem necessary to describe either the A.I.O. variables or the product usage variables in any detail, since they are not used in any subsequent analysis and a brief description of them can be found in Section 1.42. However, the demographic variables need further definition.

Table 6.1 presents a variable dictionary for the demographic factors and their corresponding scoring ranges. Note all of the variables are of a discrete nature. Three of the variables--home ownership, length of residence and location of previous dwelling are binary. A breakdown of the occupation, education, income and family life cycle categories is shown in Table 6.2.

6.3 - VARIABLE CODING AND POPULATION CHARACTERISTICS

Since it was the purpose here to apply multinomial classification procedures to the data, six out of the nine demographic factors had to be dichotomized. Common practice in analysis is to either use the numeric scores assigned to each level of the variable, or to dichotomize each variable so as to produce dummy variables. With the dummy variable approach, there are N-1 dummy variables for each factor where N is the number of levels of the factor. It was decided not to use this approach except on marital status since the remaining variables (number of rooms, location of previous dwelling, head of household's education, head of household's occupation, family income, and family life cycle) have far

TABLE 6.1

VARIABLE DICTIONARY

Variable Number	Description	Scoring	Range
1	Home ownership	1 (own)	-0 (rent)
2	Number of rooms	1 (one room)	-10 (ten or more rooms)
3	Length of residence	1 (more than five years)	-0 (five years or less)
4	Location of previous dwelling	1 (same town)	-5 (outside U.S.A.)
5	Marital status	1 (married) -2(single)	-3(widowed, separated or divorced)
6	Head of household's occupation	1 (professional)	-10 (laborer)
7	Head of household's education	1 (some grade school)	-9 (Master's or doctorate)
8	Family income	1 (under $3,000)	-10 (over $25,000)
9	Family life cycle	1 (head of household) less than 55 years, no children	-7 (head of household) 55 years or older, unemployed

TABLE 6.2

BREAKDOWN OF OCCUPATION, EDUCATION, FAMILY INCOME, AND FAMILY LIFE CYCLE CATEGORIES

I. Head of Household's Occupation Categories:

Code	Description
1	Professional, technical
2	Manager, official, proprietor
3	Sales, clerical worker
4	Craftsman, foreman
5	Operator
6	Service worker
7	Housewife
8	Student
9	Retired
10	Laborer

II. Head of Household's Education Categories:

1	Some grade school
2	Grade school completed
3	Some high school
4	High school completed
5	Some college
6	College graduate
7	Some graduate work
8	Master's or Doctorate degree

III. Family Income Categories

1	Under $3,000
2	$3,001-$5,000
3	$5,001-$7,000
4	$7,001-$9,000
5	$9,001-$11,000
6	$11,001-$13,000
7	$13,001-$15,000
8	$15,001-$20,000
9	$20,001-$25,000
10	Over $25,000

Table 6.2 (continued)

IV. Family Life Cycle Categories:

Code	Description
1	Head of household less than 55 years old, single (widowed, separated or divorced) no children.
2	Head of household less than 55 years old, married, no children.
3	Head of household less than 55 years old, with children (none teenagers).
4	Head of household less than 55 years old, with children (at least one teenager).
6	Head of household at least 55 years old, employed.
7	Head of household at least 55 years old, unemployed.

too many categories. To employ a dummy variable approach on these factors would increase the dimensionality of the problem to an unmanageable level since the total number of variables would increase to 49.

Hence, a procedure to dichotomize the remaining variables was developed. To illustrate, let x_j^* represent the cut-off point where the difference between the relative cumulative frequencies for each population is a maximum. Now, let

$F_i(xj)$ = percent of households in population i with j^{th} variable less than or equal to x_j,

then the cut-off point x_j^* is given by

$$x_j^* = \underset{x}{\text{Max}} \; |F_1(x)-F_2(x)|$$

Choosing x_j^* in this way maximizes the difference between the marginal probabilities $(p_{2j}-p_{1j})$ in the two populations. A revised description of the coding of each demographic variable appears in Table 6.3. Note, originally variable 5 (marital status) was coded by the method of dummy variables (since it had three levels): married or not, and single or not. However, a study of the marginal distributions and intercorrelations of these variables led to the conclusion that they were tapping similar dimensions. For example, high negative intercorrelations were found for these variables across both populations. Married or not, and single or not, had a correlation of -.86 in population 1, and -.76 in population 2. Hence, it was decided to dichotomize

TABLE 6.3

VARIABLE IDENTIFICATION

Variable Number	Description	Code 1 Meaning	Code 0 Meaning
1	Home ownership	Own	Rent
2	Number of rooms	More than five rooms	Five rooms or less
3	Length of residence	More than five years	Five yers or less
4	Location of previous dwelling	Outside county, state or U.S.A.	Within same county
5	Marital status	Married	Single, widowed, separated or divorced
6	Head of household's occupation	Professional, manager or clerical worker	Craftsman, operator, worker, housewife, student, retired, laborer
7	Head of household's education	Some college or above	No college
8	Family income	$11,000 or above	Less than $11,000
9	Family life cycle	55 years or older, employed or unemployed	Under 55 years old

marital status as one variable, as described in Table 6.3.

The marginal distribution of each variable and the correlations among the variables in each population are contained in Tables 6.4 and 6.5. It is apparent that several of the correlations among the variables would have to be considered significant. In fact, for the nine variables under examination, having 36 intercorrelations, a little over one-third of them (in each population) were significant at the .001 level. The implication of these significant correlations will be discussed in the next section.

6.4 - DISCRIMINATION OF HEAVY AND LIGHT PRODUCT USERS

The results of the Monte Carlo sampling described in the previous chapter can now be tested on this data. However, before evaluating the performance of each discrimination procedure with respect to the problem at hand, two points need further clarification.

First, an estimate of the a priori probability of group membership, Π, had to be determined. In the absence of additional information on the usage rates for the total universe of households serviced by this company, the best estimate is the ratio of households whose usage rates were more than \$5.00 to the total number of households included in the study. The ratio is quite close to .504 and it is more or less constant over the three areas sampled (see below).

TABLE 6.4

MARGINAL DISTRIBUTIONS

Frequency (percent) Coded 1

Variable Number	Description	Population 1 (Usage $5, n=230)		Population 2 (Usage $5, n=230)	
1	Home ownership	200	(85.5)	187	(81.3)
2	Number of rooms	167	(71.4)	135	(58.7)
3	Length of residence	132	(56.4)	111	(48.3)
4	Location of previous dwelling	54	(23.1)	28	(12.2)
5	Marital status	203	(86.8)	190	(82.6)
6	Head of household's occupation	150	(64.1)	100	(43.5)
7	Head of household's education	159	(67.9)	124	(53.9)
8	Family income	192	(82.1)	132	(57.4)
9	Family life cycle	47	(20.1)	64	(27.8)

TABLE 6.5

VARIABLE INTERCORRELATIONS BY POPULATION

Population 1 Variable

VARIABLE	1	2	3	4	5	6	7	8	9
1	1								
2	.49	1							
3	-.22	-.14	1						
4	-.12	-.01	.48	1					
5	.23	.20	.01	.09	1				
6	.07	.22	.02	.07	.13	1			
7	.08	.17	.08	.18	.06	.46	1		
8	.22	.30	-.05	.04	.18	.25	.25	1	
9	.02	.01	-.35	-.15	.01	-.18	-.09	-.15	1

Population 2 Variable

VARIABLE	1	2	3	4	5	6	7	8	9
1	1								
2	.39	1							
3	-.27	-.16	1						
4	-.05	.02	.39	1					
5	.32	.24	.05	.03	1				
6	.09	.25	.07	.13	.12	1			
7	.10	.18	.09	.13	.13	.42	1		
8	.16	.35	-.03	.05	.30	.40	.26	1	
9	-.01	-.09	-.33	-.11	-.36	-.17	-.07	-.25	1

Area	Total Number of Households	Number of Households with usage rates $\geq$\$5	Proportion of Households with usage rates $\geq$ \$5
Indianapolis	158	73	.462
Atlanta	152	80	.526
Los Angeles	154	81	.526
	464	234	.5043

Therefore, Π is set to .504 and the critical value

$$c = \log_e \, [\Pi/(1-\Pi)]$$

is .016. Hence, the Bayes decision rule for a discrimination procedure, denoted by [*] is given by

$$\hat{\beta}(x;[*]) = \begin{cases} 0 & \text{if } \hat{L}(x;[*]) < .016 \\ .496 & \text{if } \hat{L}(x;[*]) = .016 \\ 1 & \text{if } \hat{L}(x;[*]) > .016 \end{cases}$$

where $\hat{L}(x;[*]) = \log_e(\hat{\pi}_2(x;[*])/\hat{\pi}_1(x;[*]))$.

The second point to be made concerns the method for evaluating each discrimination procedure. Ideally, a split-sample method should be utilized to evaluate the performance of any classification rule with care being taken in deciding how to split the original sample. However, one of the major limitations of this method is that it requires rather large initial samples. (A complete discussion of the limitations of this method was presented in Section 3.3. This problem is particularly acute for multinomial classification procedures. To illustrate, consider the sample sizes in this study. Here, n_1=234 and n_2=230. However, with nine variables, there are 2^9=512 possible response patterns. To split the samples into smaller sub-groups seems extremely impractical since it would mean that a far greater number of states would be empty and hence the estimates are likely to be even more unstable. Thus, under these circumstances, it is doubtful that the rules $\hat{\beta}(x;[*])$ used to classify the remaining members of the hold-out sample would be very good.

For these reasons, a split-sample validation method was not utilized. Therefore, in subsequent analysis, each discrimination procedure is evaluated in terms of its apparent error. Although a bias is introduced when the same data are used to estimate a procedure and evaluate its performance, this does not, however, prevent a comparison of the relative performance of each procedure, nor preclude statements as to whether the multinomial procedures offer additional insights into the utilization of demographic factors under analysis.

The remainder of this section is divided into two parts. The first part presents the results of discrimination, using all nine demographic variables, while the second part considers discrimination with a reduced subset of variables.

6.41 - Discrimination Using All Nine Variables: Examination of the correlations among the variables, given in Table 6.5, shows that, for the most part, they cannot be considered small. Both populations contain relatively high positive and negative correlation terms. Based on this observation, the first order independent and the LDF procedures are not expected to perform well and to yield the highest apparent errors.

With nine variables there are a total of 2^9 or 512 response patterns. With so many possible response patterns it is impossible for all patterns to occur in samples of size 234 and 230. Thus, a large number of the estimates of the likelihood ratios under a full multinomial model will be based on empty cells. For example, 424 and 420 of the 512 cells in population 1 and population 2 were empty, respectively. In population 1 and population 2 combined 378 of the 512, or 74 percent of the cells, were empty. The presence of empty cells, or sparse data, will always adversely affect the estimation procedure, especially for the full multinomial model; however, in this case, there is some solace in the fact that the large percentage of the total number of empty cells appeared in both populations.

Table 6.6 gives the apparent errors for each discrimination procedure.

Because a bias is introduced whenever evaluation is based on the apparent error, care must be taken when drawing conclusions. However, it seems fairly safe to conclude that the relative performance of the full multinomial, Matusita, and complete models was substantially better than the other procedures. With these procedures the error rates were all about 25 percent, whereas the use

TABLE 6.6

APPARENT ERRORS USING ALL NINE VARIABLES

Model	Probability Member of Pop.1 Mis-classified	Probability Member of Pop.2 Mis-classified	Probability Misclassification in Combined Populations
Full multinomial	.1670	.3435	.2545
First order independence	.3662	.4057	.3858
Second order	.3139	.4110	.3621
LDF	.3030	.4300	.3664
Matusita	.1923	.3178	.2545
Martin and Bradley:			
Complete	-	-	.2522
Incomplete:			
Main effects	-	-	.4536
First order interaction	-	-	.4201

of the linear models or the second order procedure yielded error rates above 36 percent.

The relatively poor performance of the second order model was somewhat surprising, particularly considering the correlation structures in the two populations. Possibly, the poor performance with this procedure was in part due to the number of times the second order approximations to densities were negative. It turned out that 101 of the 512 probabilities $\hat{\pi}_1(x;[2])$ (or 19.7 percent) and 136 of the 512 probabilities $\hat{\pi}_2(x;[2])$ (or 26.5 percent) were negative. In order to be able to assign log likelihood ratios to response patterns x where the second order estimates were negative, the value of $\hat{\pi}_i(x;[2])$ was sent to 10^{-5}.

Table 6.7 displays the coefficients for the LDF procedure and the Martin and Bradley complete model. The coefficients appearing under the LDF column are the usually standardized linear discriminant weights which provide a measure of the contribution of each variable to discrimination. The second column of coefficients corresponds to the main effects and first order interactions for the complete model; $a_{\gamma}^{(i)}$ where $a_0^{(i)}$ is a population characteristic; $a_{\gamma}^{(i)}$ corresponds to the main effect for the j^{th} variable; and $a_{j.k}^{(i)}$ $j \neq k$, corresponds to the interaction between the j^{th} and k^{th} variables. In terms of the classification problem, $a_j^{(i)}$ measures the ability of the j^{th} variable as a discriminator whereas $a_{j.k}^{(i)}$ measures the joint ability of

TABLE 6.7

SUMMARY OF COEFFICIENTS FOR THE LDF AND COMPLETE MODELS

Variable	LDF	Complete
X_0	-	-0.0142
X_1	0.0602	0.0120
X_2	0.0852	-0.0629
X_3	0.1260	-0.0195
X_4	0.2390	-0.0368
X_5	-0.0879	0.0010
X_6	0.2636	-0.0414
X_7	0.0273	-0.0121
X_8	0.5494	-0.0145
X_9	0.0078	0.0021
X_1X_2	-	-0.0281
X_1X_3	-	-0.0187
X_1X_4	-	0.0032
X_1X_5	-	0.0478
X_1X_6	-	0.0004
X_1X_7	-	0.0062
X_1X_8	-	0.0202
X_1X_9	-	-0.0084
X_2X_3	-	0.0118
X_2X_4	-	-0.0295
X_2X_5	-	0.0108
X_2X_6	-	-0.0304
X_2X_7	-	0.0276
X_2X_8	-	-0.0015
X_2X_9	-	0.0266
X_3X_4	-	-0.0042
X_3X_5	-	0.0012
X_3X_6	-	0.0026
X_3X_7	-	0.0077

Table 6.7 (continued)

Variable	LDF	Complete
X_3X_8	-	0.0013
X_3X_9	-	-0.0078
X_4X_5	-	0.0161
X_4X_6	-	-0.0344
X_4X_7	-	-0.0108
X_4X_8	-	-0.0026
X_4X_9	-	0.0005
X_5X_6	-	0.0282
X_5X_7	-	0.0069
X_5X_8	-	-0.0038
X_5X_9	-	-0.0204
X_6X_7	-	0.0151
X_6X_8	-	-0.0339
X_6X_9	-	-0.0070
X_7X_8	-	-0.0001
X_7X_9	-	0.0188
X_8X_9	-	0.0155

the j^{th} and k^{th} variables as discriminators. Note the LDF coefficients are interpreted with respect to both sign and magnitude, while the complete model coefficients are interpreted with respect to magnitude only, since the sign of each coefficient is arbitrary.

A study of the two sets of coefficients revealed somewhat different conclusions. For the LDF procedure, the first four variables, in descending order of importance, were family income (X8), head of household's occupation (X6), location of previous dwelling (X4), and length of residence (X3). For the complete model, the first four variables, in descending order of the absolute magnitude of their coefficients (main effects) corresponded to number of rooms (X2), head of household's occupation (X6) location of previous dwelling (X4), and length of residence (X3). Note, variable X8, family income, which had the largest coefficient with the LDF, ranked fifth. Although the ability of variable X1, home ownership, and variable X5, marital status, were relatively minor when considered alone, their joint ability as discriminators was considerably more important. A similar relationship was found for variables X7 and X9, head of household's education and family life cycle.

For the Martin and Bradley complete model the ten largest interactions, in descending order of magnitude, were home ownership and marital status (X1, X5), head of household's occupation and family income (X6, X8), location of previous dwelling and head of household's occupation

(X4, X6), number of rooms and head of household's occupation (X2, X6), number of rooms and location of previous dwelling (X2, X4), marital status and head of household's occupation (X5, X6), home ownership and number of rooms (X1, X2), number of rooms and head of household's education (X2, X7), number of rooms and family life cycle (X2, X9), and marital status and family life cycle (X2, X9).

Before proceeding to discuss the results for discrimination with a reduced set of variables, the performance of the Martin and Bradley incomplete models deserves some mention. It is apparent that the use of either the main effects model or the first order interaction model yield error rates which are greater than any other procedure. On the basis of this observation plus the relatively poor performance of the second order procedure, it would seem likely that higher order terms (interactions) are present in the data. Therefore, a useful procedure, before a discriminate analysis is performed, might be to examine, say, the third order mixed moments in each population, in addition to the correlations. In this way, a researcher will be better able to decide whether procedures which involve fewer parameters to estimate can yield special efficacy.

For the most part, the results presented in this section are in agreement with the conclusions reached on the basis of the Monte Carlo sampling experiments of the previous chapter.

<u>6.42 - Discrimination Using A Reduced Set of Variables</u>: Perhaps as important as the construction of reasonable classification rules, is the choice of variables, and choosing good subsets of variables for discrimination. Although the probability of misclassification cannot be decreased by decreasing the number of variables, finding a good subset of variables is particularly important in an economic sense, since frequently variables are both expensive and difficult to obtain. Determining the worth of a variable--and choosing good subsets of variables--has received some attention in the statistical literature; however, only recently has a sample-based procedure for selecting an optimal subset of variables for the two-group multinomial classification problem been developed (Goldstein and Rabinowitz, 1975). Based on this work, the following discussion will examine the selection of demographic variables for the discrimination problem at hand.

Using the notation of Goldstein and Rabinowitz (1975), let the two groups G_1 and G_2 mix in a large population with respective prior probabilities q_1 and q_2 with $q_1>0$, $q_2>0$, $q_1 + q_2 = 1$. A p-component response vector is denoted by X and the class-conditional distributions of X within the two disjoint groups G_1 and G_2 is denoted by F_1 and F_2 with respective probability mass functions f_1 and f_2. A classification rule is defined as an ordered partition of $D = \langle D_1, D_2 \rangle$ of the sample space with unconditional probability of correct classification given by $r(D)$, where

$$r(D)=E\{r(D|x)\}=\sum_{D_1} g_1(x)+\sum_{D_2} g_2(x).$$

The discriminant scores g_1 and g_2 above are defined by $g_i(x)=q_i f_i(x)$, $i=1,2$.

Goldstein and Rabinowitz (1975) seek good subsets of variables on the basis of which set yields the largest scaled value of d, where

$$d=\min|(g_1(x))^{1/2}-(g\ \ (x))^{1/2}|$$

In terms of discriminant scores this requires choosing that subset of variables satisfying

$$\max_{1\leq j\leq p}\ \max_{1\leq i\leq \binom{p}{j}}\ \min_{x(i)}\ |g_1(x_{(i)}))^{1/2}-(g_2(x_{(i)}))^{1/2}|/(\pi_j \ell_j)^{1/2}.$$

Where x(i) represents a particular combination, when j out of the available p variables are used, $j=1,2,\ldots,p$; $i=1,2,\ldots,\binom{p}{j}$. The quantity $(\pi_j \ell_j)$ represents the scaling factor determined by the dimensionality of X(i). In particular, if x is a multivariate binary vector of dimension p, then $\pi_j \ell_j=2^{p-k}$.

The sampled based analogue of the above is to choose that subset X(i) which satisfies

$$\max_{1\le j\le p}\ \max_{1\le i<\binom{p}{j}}\ \min_{x(i)} \left|(N_1(x_{(i)}))^{1/2}-(N_2(x_{(i)}))^{1/2})\right|/(\pi_j \ell_j)^{1/2}$$

where N_h, h=1,2, denote the number of sample observations from group G_h and $N_h(x_{(i)})$, the number from G_h having $X_{(i)}=x_{(i)}$. Goldstein and Rabinowitz (1975) note that reducing both frequency distributions for a given subset of variables by considering only the scaled minimum of $|N_1(x_{i}))^{1/2}-(N_2(x_{(i)}))^{1/2}|$, throws away information which becomes more acute as the sample sizes decrease. Hence, they suggest the use of

$$\max_{1\le j\le p}\ \max_{1\le i\le\binom{p}{j}}\ \operatorname*{Avg}_{x(i)} \left| N_1(x_{(i)}))^{1/2} - (N_2(x_{(i)}))^{1/2}\right|/(\pi_j \ell_j)^{1/2},$$

which maximizes the scaled average value of

$$|N_1(x_{(i)}))^{1/2}-(N_2(x_{(i)}))^{1/2}|.$$

Moreover, Goldstein and Rabinowitz (1975) note that averaging $|(N_1(x_{(i)}))^{1/2}-(N_2(x_{(i)}))^{1/2}|$ will, in general, give better results than the minimization process. The former uses more information from the frequency distributions involved. However, it is likely to yield an optimal subset containing more variables than the minimization process.

According to the minimization procedure, the subset chosen consists of only one variable, X8, family income, whereas the averaging procedure selected the set

(X2,X3,X5,X6,X7,X8,X9)

corresponding to variables: number of rooms, length of residence, marital status, head of household's occupation, head of household's education, family income, and family life cycle.

To get a rough idea of how well variable X8, family income, discriminates alone, let an individual with response pattern x be assigned to G_1 if and only if $N_1(x)>N_2(x)$. This is only approximately, since again the same data is used both for estimation and as test points. Using family income (X8) alone resulted in 174 misses, or an apparent error rate of 37.5 percent.

Summary results for discrimination using the subset

(X2,X3,X5,X6,X7,X8,X9)

are presented in Table 6.8. The use of the full, Matusita and complete models again resulted in the smallest apparent errors (of about 29.8 percent). This translates intc an increase in apparent error of approximately 4 percent over that obtained when all nine variables were included. One particularly surprising result is the somewhat lower error rates for the first, LDF and incomplete models with the reduced subset of variables. Although the probability of misclassification cannot be decreased by decreasing the number of variables for given samples, the use of a smaller subset of variables can offer better discrimination than using the entire set. A classic example was found by Rao (1949), wherein using one variable offered better discrimination than using two.

TABLE 6.8

APPARENT ERRORS USING SUBSET (X2,X3,X5,X6,X7,X8,X9)

Model	Probability Member of Pop.1 Misclassified	Probability Member of Pop.2 Misclassified	Probability Misclassification in Combined Populations
Full multinomial	.2137	.3834	.2978
First order independence	.3210	.4479	.3839
Second order	.2782	.4525	.3646
LDF	.2601	.4650	.3621
Matusita	.2357	.3610	.2978
Martin and Bradley:			
Complete	-	-	.2974
Incomplete:			
Main effects	-	-	.4272
First order interaction	-	-	.3807

6.5 - SUMMARY AND CONCLUSIONS

This chapter has presented the results of an application of the various discrimination procedures to data on communication buyer behavior. The problem was to discriminate heavy and light users of a specific communication product on the basis of demographic variables. For the most part, the results were in agreement with the conclusions reached on the basis of the Monte Carlo sampling experiments described in the previous chapter. Moreover, it would appear that the lower apparent error rates for the full, Matusita and complete models indicate special efficacy of these procedures for demographic data.

CHAPTER VII

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

In this chapter the research findings presented in this dissertation will be summarized and the implications of these findings, together with recommendations for future research will be presented.

7.1 - SUMMARY

The objective of this research study was to examine alternative methods of discrimination so as to provide marketing researchers with more efficient methods for analyzing questionnaire data. In particular, the study addressed the problems associated with the use of qualitative and/or categorical data in the context of a discrimination problem. Although the Fisher linear discriminant function has been the most frequently used classification procedure in marketing, a review of the relevant literature indicated that, for the most part, marketing researchers have not been cognizant of the basic assumptions underlying the optimality of this procedure. Mentioned in Chapter II is the fact that in the vast majority of empirical studies using the Fisher

LDF there is no discussion of the assumptions of normality, identical variance-covariance matrices, nor any statements as to the estimation of parameters or a priori probabilities. Moreover, the marketing literature showed a marked absence of any study evaluating the performance of the Fisher LDF with data, the components of which do not satisfy the various distributional assumptions underlying its optimality.

The need for a study validating the performance of the Fisher LDF on questionnaire type data was judged to be of extreme importance, particularly considering the discrete and often classificatory nature of marketing data.

Often, either by necessity or design, the marketing researcher must handle data, the components of which are discrete and/or binary in nature. With such data, common practice before an analysis is performed has the researcher assign numeric scores to the levels of the variables and then apply a sample-based LDF which assumes a multivariate normal structure. Clearly this type of procedure must be viewed as being highly suspect, in addition to using some rather fictitious information. Several marketing researchers have been cognizant of this anomaly and have indicated that an important new direction should be the extension of discriminant analysis to those situations where the predictor variables are classificatory.

Toward this end this research study examined the relative performance of six discrimination procedures applied to binary data under a wide variety of population structures. The discrimination procedures evaluated were (1) full multinomial model, (2) first order independent model, (3) second order model, (4) Fisher LDF, (5) Matusita model, and (6) Martin and Bradley orthogonal polynomial model. Chapter IV developed each discrimination procedure and showed how each model arises from making different assumptions concerning the structure of the state probabilities. In order to evaluate the performance of each procedure, Monte Carlo sampling experiments were initiated on different population structures. A given population structure was characterized in terms of means, p_{ij}, and correlations, $r_i(jk)$, where the subscript i referred to a specific population. By using the Bahadur reparametization, Monte Carlo samples were generated from populations specified by the designated input parameter (p_{1j}, p_{2j}, $r_1(jk)$) and $r_2(jk)$. In this way, it was possible to obtain different Monte Carlo samples by simply assigning different values to the input parameters.

The sampling experiments (Chapter V) were divided into five parts, corresponding to the five general questions outlined in Chapter I. Each question was formulated so as to organize the sampling and to facilitate the examination of different aspects and properties of each discrimination procedure. Since two of the discrimination models, the Matusita and the Martin and Bradley, had never been examined, two of the five questions sought to determine whether these procedures merit more widespread attention. The remaining questions sought to determine the effects of correlation structure on the performance of each procedure, the effects of varying the magnitude of the individual marginal probabilities while maintaining a fixed mean difference, and, finally, whether certain classification procedures could effectively discriminate on the basis of disparate correlation structures rather than on mean differences.

In addition to examining the performance of each procedure under a variety of population structures, this dissertation also considered an application to data on communication buyer behavior. In particular, the application sought to determine whether the employment of alternative discrimination procedures--those which are more compatible with categorical data--offers additional insights into the utilization of demographic variables in differentiating heavy and light product users.

To discriminate heavy and light users of a specific communication product on the basis of demographic factors was thought to be particularly relevant, since several marketing researchers were quoted to the effect that demographic factors have little role or importance as determinants or even correlates of consumption behavior.

In addition, the data provided an ideal vehicle for testing the conclusions and recommendations of the Monte Carlo sampling described in Chapter V.

The data utilized in this chapter was provided by a major United States corporation. In order to draw the total sample from the population served by a number of associated companies, individuals residing in both the central city and surrounding suburban communities were selected from three areas considered representative. The areas of sampling were Indianapolis, Atlanta and Los Angeles. An evenly dispersed quota sample based upon respondents' estimates of their product usage rates was chosen as the method of selection. Interviewers were instructed to select a number of central city and suburban communities within each of the sampling areas, comprising a mix of lower, middle and upper socio-economic households. All respondents were screened with respect to (1) their use of a specific communication product, and (2) their willingness to complete the questionnaire and have the interviewer return to pick it up. A final usable data base of 464 respondents was secured. This constituted a response rate of over 90 percent of those who agreed to complete the questionnaire.

Respondents were asked to provide information relating to three broad areas: (1) activities, interests, and opinions (A.I.O.); (2) product usage behavior; and (3) demography and socio-economic data. The demographic data utilized in this research study were: (1) home ownership, (2) number of rooms, (3) length of residence, (4) location of previous dwelling, (5) marital status, (6) head of household's occupation, (7) head of household's education, (8) family income and (9) family life cycle.

7.2 - CONCLUSIONS AND IMPLICATIONS

This section is divided into two parts. First, the major findings of the Monte Carlo sampling experiments (Chapter V) together with the results of the application of the various discrimination procedures to data on communication buyer behavior (Chapter VI) are summarized. Second, the implications of these findings and recommendations are presented.

7.21 - Summary of Research Findings:

The following summarizes the major research findings of the series of Monte Carlo sampling experiments described in Chapter V:

1. It was always possible to determine critical bounds on the correlations such that, if they are exceeded, the use of the first order independent model or the LDF will result in greater misclassification error (on the average) than either the full multinomial, second order, or Matusita procedures. In addition, for population structures containing small mean differences, the critical bounds are likely to be more restricted.

2. The evidence suggests that highly correlated variables may yield better discrimination than uncorrelated variables. Also, for fixed mean differences, negative correlations were found to improve discrimination across all procedures. That is, with negative correlations the optimum error and the theoretical errors were lower than with positive correlations.

3. The linear models appeared most sensitive to large mean differences. With large mean differences, the performance of both the first and LDF procedures substantially improved, and their performance was not significantly different from the other procedures except in those populations containing reversals. Here, the performance of the linear models was not improved by the presence of larger mean differences.

4. Most importantly, the sampling experiments revealed that the use of linear models can result in severe anomalies even when the covariance matrices in the two populations are taken to be identical, which is an underlying assumption used to derive the LDF. The true L.L.R.'s were shown to be capable of reversing both sign and direction and, hence, the linear models cannot satisfactorily characterize populations of this kind since their L.L.R.'s are always monotone.

5. For a fixed mean difference, the relative performance of each discrimination procedure did not appreciably

change as the magnitude of the marginal probabilities was varied; however, the magnitude of the resultant error rates did. Uniformly, larger optimum errors were found across all population structures as the magnitude of the individual marginal probabilities increased.

6. It was apparent that the full multinomial, second order and Matusita procedures could better utilize the information provided by disparate correlation structures than either of the linear models. For the most part, these procedures showed a marked ability to discriminate even when the mean differences were quite small.

7. An examination of the Martin and Bradley models showed the behavior of the complete model to be very similar to that of the full multinomial and Matusita procedures. The performance of the incomplete models, main effects, and first order interaction was somewhat more difficult to characterize in terms of means and correlations. Nevertheless, it appears that the use of these models on populations containing moderate correlations will yield satisfactory results. In addition, the sampling experiments demonstrated that a reduced model should not be used unless an iterative procedure is employed.

8. The performance of the Matusita model with unequal sample sizes appeared quite satisfactory with respect to both the mean apparent error and the mean actual error. Even when the sample size in one population was twice that

of the other, the Matusita model and the full multinomial model gave near identical results across most population structures. In addition, the results suggest that the Matusita model may be particularly useful on population structures containing extreme correlations or when data are sparse.

In turning to the application of these discrimination procedures to data on communication buyer behavior where the problem was to discriminate heavy and light users of a specific communication product on the basis of demographic factors, the major findings were:

1. The use of the full multinomial, Matusita and complete models resulted in apparent errors which were considerably lower than those for the other procedures. With these procedures, the error rates were all about 25 percent, whereas the use of either the first order independent, second order, LDF, or incomplete procedures yielded error rates above 36 percent.

2. The relatively poor performance of the second order model was somewhat surprising, particularly considering the number of significant correlation terms present. Possibly, the poor showing of this procedure was in part due to the number of times the second order approximations to densities were negative.

3. The LDF procedure revealed the following four variables to be most important in discriminating heavy and

light users (in descending order of importance): family income, head of household's occupation, location of previous dwelling, and length of residence. For the complete model, the first four variables, in descending order of the magnitude of their main effects, were: number of rooms, head of household's occupation, location of previous dwelling, and length of residence.

4. With respect to the interactions among variables, the complete model revealed several variables which had relatively small main effects, but relatively large interactions. In other words, there were several variables which had relatively little ability to discriminate when considered alone; however, when considered jointly with another variable their contribution to discrimination was considerably greater. Two pairs of such variables were home ownership and marital status, and head of household's education and family life cycle.

5. The variable selection procedure developed by Goldstein and Rabinowitz (1975) resulted in two subsets. The minimization procedure selected only one variable--family income--whereas the averaging procedure selected the set corresponding to the variables: number of rooms, length of residence, marital status, head of household's occupation, head of household's education, family income, and family life cycle. Discrimination with family income alone yielded an apparent error of approximately 37.5 percent, while using the set of seven variables resulted in an apparent error rate of about 29.8 percent.

7.22 - Conclusions and Recommendations: The research findings described above lead to the following conclusions and general recommendations:

1. The linear models (first and LDF) should not be used whenever the mean vectors in the two populations are similar and when it is suspected, or known beforehand, that the correlations are large. Even when the samples are taken from populations which have identical variance-covariance matrices, the use of the linear model is likely to yield significantly greater misclassification (on the average) than either the full multinomial, second order, or Matusita procedures.

2. For populations containing relatively high correlations, the use of either the full multinomial, second order, Matusita, or complete models is recommended. The choice of one of these procedures will depend on, among other things, the degree of sparseness of the data and the presence of higher order interactions. With higher order interactions and little sparseness, the use of a full multinomial or a complete model will probably yield "good" results. On the other hand, the Matusita procedure appears potentially useful on data which are sparse.

3. Generally, it is fairly safe to conclude that with large mean differences most discrimination procedures do well and hence the selection of a particular procedure is a far less serious problem. Moreover, with large mean

differences and moderate correlations, the researcher can utilize those procedures which involve fewer parameter estimates.

4. The Martin and Bradley models offer the researcher the ability to identify the joint contribution of variables to discrimination. Given this rather unique feature and the fact that the performance of the complete model is nearly identical to that of the full multinomial model under a variety of population structures, it would seem that marketing researchers can gain additional insights into the relationship among variables by utilizing this procedure. For example, in applying the complete model to the data on communication buyer behavior it was possible to talk to relationships among the demographic variables not readily accessible when using the other procedures.

5. Perhaps most importantly, this dissertation has demonstrated the need for marketing researchers to examine the underlying structure of the data before an analysis is performed. A useful procedure, therefore, before a discrimination analysis is performed, might be to examine, say, the third order mixed moments in each population in addition to the correlations so as to be better able to determine whether procedures which involve fewer parameter estimates will yield satisfactory results.

6. The key to the effective use of demographic factors would appear to lie in the application of statistical

techniques which better accommodate the data. Clearly, to use statistical procedures which impose assumptions of normality, linearity, and additivity on the relationship among the variables seems overly restrictive and self-defeating. In particular, this study has demonstrated the special efficacy of alternative discrimination procedures in utilizing the information provided by demographic factors in discriminating heavy and light users of a communication product.

7.3 - RECOMMENDATIONS FOR FURTHER RESEARCH

One of the main objectives of this dissertation was to provide a foundation for further research in the area of classification problems. Several areas which were not extensively addressed in this study warrant further investigation. Among these areas are:

1. The extension of the Bahadur parametric representation of the joint distribution of M-dichotomous variables to the case where each X_j can assume more than two different values.

2. An area in need of further research concerns the problems arising from sparse data. Sparse data may seriously affect the estimation of parameters and hence techniques for overcoming this problem need further development. In particular, the Matusita model seems potentially useful in this area and hence merits more widespread attention.

3. Of major importance is the need to develop discrimination procedures which can accommodate mixtures of both continuous and discrete variables. Procedures of this kind will be better able to incorporate all of the information provided by the variables into the classification rule. An attempt at constructing such discriminant functions has been recently reported by Krzanowski (1975).

4. Finally, there is a real need for marketing researchers to reexamine the role and importance of demography as determinants of consumption behavior. Toward this end, it might be of interest to investigate other discrete multivariate techniques, such as multidimensional contingency models, to see whether these kinds of procedures uncover new insights into the use and interpretation of demographic information.

APPENDICES

APPENDIX I

APPENDIX I

THE PARAMETRIC REPRESENTATION FOR THE JOINT DISTRIBUTION OF M- DICHOTOMOUS VARIABLES

This appendix contains a brief description of the Bahadur reparametization (1961) together with the restrictions on p_{ij} and $r_i(jk)$ necessary to ensure that the Second order model leads to a probability distribution.

Let $X = (X_1, X_2, \ldots, X_m)$ be an m- dimensional random vector of Bernoulli variables. A particular value of X, denoted by $x = (x_1, x_2, \ldots, x_m)$ is called a response pattern. The probability that response pattern x is observed in the i^{th} population (i = 1,2) will be denoted by $\pi_i(x)$. Now, Bahadur (1961) has shown that the multinomial probabilities $\pi_i(x)$ can be reparametized as follows. Define the standardized random variable Z_{ij} by

$$Z_{ij} = (X_j - p_{ij}) / (p_{ij} (1 - p_{ij}))^{1/2},$$

where

$$p_{ij} = E_i(X_j), \text{ the expectation of } X_j \text{ in population } i$$
$$(i = 1,2;\ j = 1,2, \ldots > m).$$

Let

$$r_i(jk) = E(Z_{ij}\ Z_{ik})$$
$$r_i(jkl) = E(Z_{ij}\ Z_{ik}\ Z_{il})$$

.

.

.

.

$$r_i(12 \ldots m) = E(Z_{i1} Z_{i2} \cdot \cdot \cdot \cdot Z_{im}),$$

then

$$\pi_i(x) = \prod_{i=1}^{m} p_{ij}^{\,xj} (1 - p_{ij})^{1-xj}$$
$$\cdot\{1 + \sum_{j<k} r_i(jk)\, Z_{ij}\, Z_{ik} + \sum_{j<k<l} r_i(jkl)$$
$$Z_{ij}\, Z_{ik}\, Z_{il} + \ldots + r_i(12\ldots m)$$
$$Z_{i1}\, Z_{i2} \ldots Z_{im}\} \qquad \text{(A1.1)}$$

Note that $r_i(jk)$ is equal to the ordinary product-moment correlation between variables X_j and X_k in population i. A proof of this is given by Bahadur (1961).

The appeal of this reparametization lies in its ability to describe the state probabilities $\pi_i(x)$ in terms of means, p_{ij}, and correlations, $r_i(jk)$. For example, the first order model was derived by ignoring all correlation terms $r_i(jk)$, $r_i(jkl)$, . . . , $r_i(12 \ldots m)$. This is equivalent to assuming independence among the variables. On the other hand, the Second order model was obtained by including only terms of the kind $r_i(jk)$. That is, all higher order correlations, $r_i(jkl)$, . . . , $r_i(12 \ldots m)$ were omitted.

A basic limitation of the Second order model is that it may lead to negative estimates for $\pi_i(x)$. Letting $\pi_i(x;(2))$ denote the second order approximation to $\pi_i(x)$, then the following proof given in Bahadur (1961) determines the conditions under which $0 \leq \pi_i(x;(2)) \leq 1$ for all patterns of x.

There is no need to retain the "i" subscript since the proof does not depend on whether the observations are in population 1 or population 2.

Let R denote the MxM correlation matrix with $r_{jj} = 1$ and second-order correlations terms $r(jk)$. Let λ_{min} denote the smallest characteristic rool of R, i.e., the smallest characteristic root of the determinantal equation $|R - \lambda I| = 0$. For each j, let $\beta_j = \max(p_j/(1 - p_j),\ (1 - p_j)/p_j)$, then $0 \leq \pi_i(x;\{2\}) \leq 1$ for all patterns x as long as

$$\lambda_{min} \geq 1 - \frac{2}{\sum_{j=1}^{m} \beta_j} . \tag{A1.2}$$

In addition, Bahadur (1961) gives the formula for finding the maximum and minimum values for the correlation between the variables in the special case where all $p_j = p$ for each j and all $r(jk) = r$ for all $j \neq k$. In this case, $0 \leq \pi_i(x;\{2\}) \leq 1$ for all x if and only if

$$- \frac{2}{m(m-1)} \cdot \min \left(\frac{p}{1-p} , \frac{1-p}{p}\right) \leq r \leq \frac{2p\ (1 - p)}{(m-1)p(1-p) + \frac{1}{4} - \gamma_o} \tag{A1.3}$$

where

$$\gamma_o = \min_{y=1,2,..,m} \{y - (m-1)p - \tfrac{1}{2}\}^2 \leq \frac{1}{4} .$$

APPENDIX II

COMPUTER PROGRAM

```
C THIS PROGRAM COMPUTES VARIOUS STATISTICS FOR THE FULL
C MULTINOMIAL,FIRST ORDER INDEPENDENT,SECOND ORDER
C FISHER LDF, MATUSITA, AND BAHADUR AND BRADLEY
C ORTHOGONAL POLYNOMIAL CLASSIFICATION MODELS.
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C P1=VECTOR OF MARGINAL PROB. FOR POP.1                 C
C P2=VECTOR OF MARGINAL PROB. FOR POP.2                 C
C RT1=NXN MATRIX OF CORRELATIONS FOR POP.1              C
C RT2=NXN MATRIX OF CORRELATIONS FOR POP.2              C
C TP1=STATE PROB. FOR POP.1                             C
C TP2=STATE PROB. FOR POP.2                             C
C NF1=OBSERVED FREQ. FOR POP.1                          C
C NF2=OBSERVED FREQ. FOR POP.2                          C
C NV=N= THE NUMBER OF VARIABLES                         C
C NT=NUMBER OF MONTE CARLO TRIALS                       C
C NS1=SAMPLE SIZE FOR POP.1                             C
C NS2=SAMPLE SIZE FOR POP.2                             C
C IX=SEED FOR RANDOM SAMPLING.                          C
C INDEX=1 IF MARGINAL PROB. ARE SPECIFIED               C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
// JOB TIME=10
//*MAIN LINES=30
// EXEC FORTGCLG,REGION.GO=130K
//FORT.SYSIN DD *
      DIMENSION TP1(64),TP2(64),XLR(5,64),EP(3,5),PF1(64),
     1PF2(64),NF1(64),NF2(64),APER(3,5),ACER(3,5),SLR(5,64),
     2COR(5),ACOR(5),P1(6),P2(6),RT1(6,6),RT2(6,6),ERRINC(5)
      DIMENSION XNC1(2),XMP1(2),XMC2(2),XMM2(2),R1(2),R2(2)
      DIMENSION FF(10),TT(3,5)
      DIMENSION XMP1(64),XMP2(64),XNF1(64),XNF2(64)
    1 READ(5,100)NV,NT,NS1,NS2,IX,INDEX
      IF(NV.EQ.0)GO TO 99
      WRITE(6,101)NV,NT,NS1,NS2,IX,INDEX
  101 FORMAT(1H1,' INPUT VALUES ARE NV=',I5,' NT=',I5,' NS=',2I5,
     1' IX=',I5,' INDEX=',I3//)
      LL=6
      PI=0.5
      NC=2**NV
      XNS1=NS1
      XNS2=NS2
      XNT=NT
      IF(INDEX.EQ.1)GO TO 2
      READ(5,102)TP1
      READ(5,102)TP2
      GO TO 3
    2 READ(5,102)P1,P2
      WRITE(6,103) P1,P2
  310 FORMAT(1H ,F10.6)
  103 FORMAT(/' INPUT PROBABILITIES'/(10X6F6.2)/(10X6F6.2))
  102 FORMAT(16F5.4)
      READ(5,104)RT1
      READ(5,104)RT2
  104 FORMAT(6F5.2)
      WRITE(6,105)RT1
      WRITE(6,105)RT2
  105 FORMAT(/' INPUT CORRELATIONS'/(5X6F6.2))
  100 FORMAT(10I5)
      CALL MULTI(NV,P1,RT1,PF1,TP1,MB1)
      CALL MULTI(NV,P2,RT2,PF2,TP2,MB2)
    3 CALL PAIR(NV,TP1,TP2,XLR,EC,NS1,NS2,NP1,NP2,LL,XNS1,XNS2,NF1,NF2)
      WRITE(6,210)
  210 FORMAT(//15X,'THEORETICAL ERROR'//18X,'FULL',5X,
     1'FIRST',4X,'SECOND',7X,'LDF',2X,'MATISUTA')
      DO 400 I=1,LL
```

```
      XLR(5,J)=ER(3,J)
  440 CONTINUE
      DO 450 I=1,NC
      WRITE(6,460)(XLR(I,J),J=1,5)
  450 CONTINUE
  460 FORMAT(3X,'L.L.R.',1X,5F12.7)
      CALL XMTSS(TP1,TP2,XLR,FR,PT,NC)
      CALL CORFL(XLR,NC,XLR,COR)
      CALL DOUT1(FR,COR)
      CALL FTEST(XNS1,XNS2,FR,FF,DF1,DF2)
      WRITE(6,430)FF,DF1,DF2
  430 FORMAT(/,'F TEST',1X,1PE8.3,5X,'DF=',1X,F4.0,1X,F4.0)
      THER=ER(3,1)
      WRITE(6,203)NR1,NR2
  203 FORMAT(/'NO. OF TIMES SECOND ORDER PROBS.LT.ZERO',2I5)
      MR1=0
      MR2=0
      DO 6 J=1,5
      ACOR(J)=0.0
      DO 6 I=1,3
      APER(I,J)=0.0
    6 ACER(I,J)=0.0
      DO 20 KK=1,IT
      LL=1
      CALL RSAM(TP1,NF1,NS1,IX,NV)
      CALL RSAM(TP2,NF2,NS2,IX,NV)
      DO 11 I=1,NC
      PF1(I)=NF1(I)/XNS1
   11 PF2(I)=NF2(I)/XNS2
      CALL PXLR(NV,PF1,PF2,SLR,NC,NS1,NS2,NR1,NR2,LL,XNS1,XNS2,NF1,NF2)
      CALL CORFL(XLR,NC,SLR,COR)
      MR1=MR1+NR1
      MR2=MR2+NR2
      CALL XMTSS(PF1,PF2,SLR,FR,PT,NC)
      DO 15 I=1,3
      DO 15 J=1,5
   15 APER(I,J)=APER(I,J)+ER(I,J)/XNT
      CALL XMTSS(TP1,TP2,SLR,FR,PT,NC)
      DO 17 J=1,5
      ACOR(J)=ACOR(J)+COR(J)/XNT
      DO 17 I=1,3
   17 ACER(I,J)=ACER(I,J)+ER(I,J)/XNT
   20 CONTINUE
      WRITE(6,410)IT,NS1,NS2
  410 FORMAT(//15X'APPARENT ERROR',I5,2X'SAMPLES OF',2I5/)
      CALL DOUT1(APER,COR)
      CALL FTEST(XNS1,XNS2,APER,FF,DF1,DF2)
      WRITE(6,430)FF,DF1,DF2
      WRITE(6,420)
  420 FORMAT(//15X'ACTUAL ERROR'/)
      CALL DOUT1(ACER,COR)
      CALL FTEST(XNS1,XNS2,ACER,FF,DF1,DF2)
      WRITE(6,430)FF,DF1,DF2
      DO 30 L=1,5
   30 ERRINC(L)=100.*(ACER(3,L)-THER)
      WRITE(6,205) ERRINC
  205 FORMAT(/'INCREASE',1X5F10.2)
      WRITE(6,203) MR1,MR2
      WRITE(6,480)
  480 FORMAT(3X,'NF1',2X,'PF1',2X,'NF2',2X,'PF2')
      DO 490 I=1,NC
  490 WRITE(6,500)NF1(I),PF1(I),NF2(I),PF2(I)
  500 FORMAT(1X,I5,F5.3,I5,F5.3)
      CALL PLOT(NV,NC,NF1,NF2,NS1,NS2)
      GO TO 1
   99 CALL EXIT
```

```
      END
C PRINTS MISCLASSIFICATION PROB. FOR POP.1 AND POP.2
      SUBROUTINE POUT1(APER,COR)
      DIMENSION APER(3,5),COR(5)
      PI=0.5
      WRITE(6,211)PI
  211 FORMAT(/'PI=',F5.2)
      WRITE(6,212)(APER(1,I),I=1,5)
  212 FORMAT(3X'P(2/1)',3X5F10.4)
      WRITE(6,213)(APER(2,I),I=1,5)
  213 FORMAT(3X'P(1/2)',3X5F10.4)
      WRITE(6,214)(APER(3,I),I=1,5)
      WRITE(6,215)COR
  214 FORMAT(2X'TOTAL P',3X5F10.4)
  215 FORMAT(/2X'ILR COR',3X5F10.4)
      RETURN
      END
C CALLS VARIOUS SUBROUTINES
      SUBROUTINE DXLR(NV,TP1,TP2,XLR,NC,NS1,NS2,NR1,NR2,LL,XNS1,XNS2,
     1NF1,NF2)
      DIMENSION TP1(64),TP2(64),XLR(5,64),P1(6),P2(6),RT1(6,6),RT2(6,6),
     1PI1(64),PI2(64),PC1(64),PC2(64),TLR(64),COV1(6,6),COV2(6,6),
     2NF1(64),NF2(64),XMP1(64),XMP2(64)
      CALL DREST(NV,TP1,P1,RT1)
      CALL DREST(NV,TP2,P2,RT2)
      CALL MULTI(NV,P1,RT1,PI1,PC1,NR1)
      CALL MULTI(NV,P2,RT2,PI2,PC2,NR2)
      CALL LOGLR(TP1,TP2,TLR,NC)
      DO 10 J=1,NC
   10 XLR(1,J)=TLR(J)
      CALL LOGLR(PI1,PI2,TLR,NC)
      DO 11 J=1,NC
   11 XLR(2,J)=TLR(J)
      CALL LOGLR(PC1,PC2,TLR,NC)
      DO 12 J=1,NC
   12 XLR(3,J)=TLR(J)
      CALL CONCOR(P1,P2,RT1,RT2,COV1,COV2,NV)
      CALL YETSM(NV,NS1,NS2,COV1,COV2,P1,P2,TLR)
      DO 13 J=1,NC
   13 XLR(4,J)=TLR(J)
      IF(LL.EQ.0)GO TO 16
      CALL MAT(XNS1,XNS2,NF1,NF2,NC,XMP1,XMP2,TLR)
      DO 14 J=1,NC
   14 XLR(5,J)=TLR(J)
   16 RETURN
      END
C COMPUTES MISCLASSIFICATION PROB.
      SUBROUTINE XMISS(P1,P2,XLR,ER,PI,NC)
      DIMENSION P1(64),P2(64),XLR(5,64),ER(3,5)
   19 FORMAT(//,'XMISS RESULTS')
      QI=1.0-PI
      CRIT=ALOG(PI/QI)
   20 FORMAT(1H ,2X,F10.5,2X,F10.5,2X,F10.5)
   26 FORMAT(1H ,5(8F8.6,/))
      DO 1 I=1,3
      DO 1 J=1,5
    1 ER(I,J)=0.0
      DO 10 I=1,5
      DO 10 J=1,NC
      IF(XLR(I,J)-CRIT)5,6,7
    5 ER(2,I)=ER(2,I)+P2(J)
      GO TO 10
    6 ER(2,I)=ER(2,I)+PI*P2(J)
      ER(1,I)=ER(1,I)+QI*P1(J)
      GO TO 10
    7 ER(1,I)=ER(1,I)+P1(J)
```

```
   10 CONTINUE
      DO 15 I=1,6
   15 EP(3,I)=P1*EP(1,I)+Q1*EP(2,I)
      RETURN
      END
C COMPUTES FIRST AND SECOND ORDER BAHADUR PROBABILITIES
      SUBROUTINE MULTI(NV,P,R,PI1,PI2,NR)
      DIMENSION P(6),PI1(64),PI2(64),N(6),Q(6),R(6,6)
   25 FORMAT(//,'MULTI RESULTS')
      NR=0
      DO 1 I=1,NV
    1 Q(I)=1.-P(I)
      L=1
      DO 20 I1=1,2
      N(1)=I1-1
      DO 20 I2=1,2
      N(2)=I2-1
      DO 20 I3=1,2
      N(3)=I3-1
      DO 20 I4=1,2
      N(4)=I4-1
      DO 20 I5=1,2
      N(5)=I5-1
      DO 20 I6=1,2
      N(6)=I6-1
      PI1(L)=1.0
      DO 9 K=1,NV
      IF(N(K))7,7,8
    7 PI1(L)=PI1(L)*Q(K)
      GO TO 9
    8 PI1(L)=PI1(L)*P(K)
    9 CONTINUE
      NV1=NV-1
      STERM=1.0
      DO 10 I=1,NV1
      J1=I+1
      DO 10 J=J1,NV
      DENOM=P(I)*Q(I)*P(J)*Q(J)
      IF(DENOM.EQ.0.0)GO TO 11
      Z=(N(I)-P(I))*(N(J)-P(J))/SQRT(DENOM)
      GO TO 10
   11 Z=1.0
   10 STERM=STERM+Z*R(I,J)
      PI2(L)=PI1(L)*STERM
      IF(PI2(L).LT.0.0)NR=NR+1
      L=L+1
   20 CONTINUE
   21 FORMAT(1H ,5X,F5.4,2X,F5.4)
   22 FORMAT(1H ,2X,F15.7,2X,F15.7)
   26 FORMAT(1H ,2X,I5)
      RETURN
      END
C GENERATES MONTE CARLO SAMPLES
      SUBROUTINE PSAM(PF,NF,N,IX,NR)
      DIMENSION PF(64),NF(64),CUT(64)
      JJ=2**NR
      NF(1)=0
      CUT(1)=PF(1)
      DO 1 J=2,JJ
      J1=J-1
      CUT(J)=CUT(J1)+PF(J)
    1 NF(J)=0
      DO 6 K=1,N
      CALL RANDU(IX,IY,Y)
      IX=IY
      DO 5 I=1,JJ
```

```
      IF(X-CUT(I))3,3,5
    3 NF(J)=NF(J)+1
      GO TO 6
    5 CONTINUE
    6 CONTINUE
      RETURN
      END
C COMPUTES SAMPLE CORRELATIONS
      SUBROUTINE DHEST(NR,PF,P,R)
      DIMENSION PF(64),P(6),R(6,6),N(6)
      DO 2 I=1,NR
    1 P(I)=0.0
      DO 2 J=1,NR
    2 R(I,J)=0.0
   23 FORMAT(1H ,F15.7)
      L=1
      DO 10 I1=1,2
      N(1)=I1-1
      DO 10 I2=1,2
      N(2)=I2-1
      DO 10 I3=1,2
      N(3)=I3-1
      DO 10 I4=1,2
      N(4)=I4-1
      DO 10 I5=1,2
      N(5)=I5-1
      DO 10 I6=1,2
      N(6)=I6-1
      DO 5 I=1,NR
      P(I)=N(I)*PF(L)+P(I)
      DO 5 J=1,NR
      R(I,J)=N(I)*N(J)*PF(L)+R(I,J)
    5 CONTINUE
   10 L=L+1
      DO 12 I=1,NR
      DO 12 J=1,NR
      DENOM=P(I)*(1.0-P(I))*P(J)*(1.0-P(J))
      IF(DENOM.EQ.0.0)GO TO 11
      R(I,J)=(R(I,J)-P(I)*P(J))/SQRT(DENOM)
      GO TO 12
   11 R(I,J)=1.0
   12 CONTINUE
   22 FORMAT(1H ,'R=',F15.7,2X,6(F10.5))
      RETURN
      END
      SUBROUTINE LOGLR(X1,X2,XLR,K)
C THIS SUBROUTINE CALCULATES THE LOG LIKELIHOOD RATIOS
      DIMENSION X1(64),X2(64),XLR(64)
   10 FORMAT(//,2X,'LOG LIKELIHOOD RATIO RESULTS')
      DO 5 I=1,K
      IF(X1(I))2,2,3
    2 X1(I)=.000001
    3 IF(X2(I))4,4,5
    4 X2(I)=.000001
    5 XLR(I)=ALOG(X2(I)/X1(I))
    6 FORMAT(1H ,5X,F15.7,2X,F15.7)
    7 FORMAT(1H ,'XLR=',F15.7)
      RETURN
      END
      SUBROUTINE XFISH(NR,NS1,NS2,COV1,COV2,P1,P2,XSCOR)
C THIS SUBROUTINE COMPUTES FISHER'S LDF IN THE SPECIAL
C CASE WHERE X(I) ARE DICHOTOMOUS
C NR=NO. OF VARIABLES
C NSI=SAMPLE SIZE FROM POPULATION I
C COVI=COVAR MATRIX FROM POPULATION I
C PI=MARGINAL PROB. VECTOR FROM POPULATION I
```

```
C XSCOR=OUTPUT VECTOR OF LD SCORES
      DIMENSION COV1(6,6),COV2(6,6),P1(6),P2(6),RLAM(6),XSCOR(64)
      DIMENSION D(6),S(6,6),LV(6),MV(6),N(6)
      XNS=XS1+XS2
      PI1=XS1/XNS
      PI2=XS2/XNS
      DO 10 I=1,NP
   10 D(I)=P2(I)-P1(I)
      DO 12 I=1,NP
      DO 12 J=1,NP
   12 S(I,J)=PI1*COV1(I,J)+PI2*COV2(I,J)
      CALL MINV(S,NP,DET,LV,MV)
      DO 15 K=1,NP
      RLAM(K)=0.0
      DO 15 I=1,NP
   15 RLAM(K)=S(K,I)*D(I)+RLAM(K)
      CONST=0.0
      DO 16 K=1,NP
   16 CONST=CONST+RLAM(K)*(P1(K)+P2(K))
      CONST=0.5*CONST
      L=1
      DO 20 I1=1,2
      N(1)=I1-1
      DO 20 I2=1,2
      N(2)=I2-1
      DO 20 I3=1,2
      N(3)=I3-1
      DO 20 I4=1,2
      N(4)=I4-1
      DO 20 I5=1,2
      N(5)=I5-1
      DO 20 I6=1,2
      N(6)=I6-1
      XSCOR(L)=0.0
      DO 18 K=1,NP
   18 XSCOR(L)=RLAM(K)*N(K)+XSCOR(L)
      XSCOR(L)=XSCOR(L)-CONST
   20 L=L+1
      RETURN
      END
C COMPUTES SAMPLE COVARIANCES
      SUBROUTINE CONCOR(P1,P2,RT1,RT2,COV1,COV2,NV)
      DIMENSION P1(6),P2(6),RT1(6,6),RT2(6,6),COV1(6,6),COV2(6,6)
      DO 5 I=1,NV
      DO 5 J=1,NV
      IF(I-J)2,3,4
    2 COV1(I,J)=RT1(I,J)*SQRT(P1(I)*P1(J)*(1.0-P1(I))*(1.0-P1(J)))
      COV2(I,J)=RT2(I,J)*SQRT(P2(I)*P2(J)*(1.0-P2(I))*(1.0-P2(J)))
      GO TO 5
    3 COV1(I,I)=P1(I)*(1.0-P1(I))
      COV2(I,I)=P2(I)*(1.0-P2(I))
      GO TO 5
    4 COV1(I,J)=COV1(J,I)
      COV2(I,J)=COV2(J,I)
    5 CONTINUE
      RETURN
      END
C COMPUTES CORRELATION BETWEEN OBS. L.L.R. AND TRUE L.L.R.
      SUBROUTINE CORREL(XLR,NC,SLR,COR)
      DIMENSION XLR(5,64),SLR(5,64),COR(5)
      SUMX=0.0
      SUMX2=0.0
      DO 5 J=1,NC
      SUMX=SUMX+XLR(1,J)
    5 SUMX2=SUMX2+XLR(1,J)*XLR(1,J)
      XBAR=SUMX/NC
```

```
      XVAR=SUMX2-XBAR*SUMX
      DO 20 I=1,5
      SUMY=0.0
      SUMY2=0.0
      SUMXY=0.0
      DO 10 J=1,NC
      SUMY=SUMY+SLR(I,J)
      SUMY2=SUMY2+SLR(I,J)*SLR(I,J)
   10 SUMXY=SUMXY+XLR(I,J)*SLR(I,J)
      YBAR=SUMY/NC
      YVAR=SUMY2-YBAR*SUMY
      COR(I)=0.0
      IF(XVAR.NE.0.0.AND.YVAR.NE.0.0)COR(I)=(SUMXY-XBAR*SUMY)/
     1SQRT(XVAR*YVAR)
   20 CONTINUE
      RETURN
      END
C COMPUTES F TEST-GOLDSTEIN(1975)
      SUBROUTINE FTEST(XNN,XMM,TT,FF,DF1,DF2)
      DIMENSION XNC1(2),XNM1(2),XMC2(2),XMM2(2),R1(2),R2(2)
      DIMENSION FF(10),TT(3,5)
      J=1
      XM=2
      DO 100 K=1,4
      KK=K+1
      DO 90 K1=KK,5
      TOT1=0.0
      TOT2=0.0
      R1(1)=0.0
      R2(1)=0.0
      R1(2)=0.0
      R2(2)=0.0
      XNC1(1)=0.0
      XNC1(2)=0.0
      XNM1(1)=0.0
      XNM1(2)=0.0
      XMC2(1)=0.0
      XMC2(2)=0.0
      XMM2(1)=0.0
      XMM2(2)=0.0
      DO 11 I=1,2
      XNM1(I)=TT(I,K)*XNN
      XNC1(I)=XNN-XNM1(I)
      XMM2(I)=TT(I,K1)*XMM
      XMC2(I)=XMM-XMM2(I)
      T1=((XM**2)*XNC1(I))-XNN
      IF(T1)16,17,16
   16 R1(I)=(XM-1)*(T1**2)*((XM**2)*XNM1(I)-(XNN*(XM-1)))**2
   17 TOT1=TOT1+R1(I)
      T2=((XM**2)*XMC2(I))-XMM
      IF(T2)15,18,15
   15 R2(I)=(XM-1)*(T2**2)*((XM**2)*XMM2(I)-(XMM*(XM-1)))**2
   18 TOT2=TOT2+R2(I)
   11 CONTINUE
      FF(J)=(XMM*TOT1)/(XNN*TOT2)
      J=J+1
   90 CONTINUE
  100 CONTINUE
      DF1=(2*XM)-1
      DF2=(2*XM)-1
      RETURN
      END
C COMPUTES SAMPLE-BASED MATUSITA MODEL
      SUBROUTINE MAT(XNS1,XNS2,NF1,NF2,N,XMP1,XMP2,TLP)
      DIMENSION XMP1(60),XMP2(60),NF1(60),NF2(60),
     1XPF1(60),XPF2(60),TLP(60)
```

```
      DO 9 I=1,K
      XNF1(I)=NF1(I)
      IF(XNF1(I).LT.2.0)XNF1(I)=2.0
      XNF2(I)=NF2(I)
      IF(XNF2(I).LT.0.0)XNF2(I)=0.2
    9 CONTINUE
      DO 10 K=1,N
      XMP1(K)=0.0
      XMP2(K)=0.0
   10 CONTINUE
      DO 100 J=1,K
      XMP1(J)=SQRT(((XNF1(J)+1)/(XNS1+1)))*(XNF2(J)/XNS2))
      XMP2(J)=SQRT((XNF1(J)/XNS1)*((XNF2(J)+1)/(XNS2+1)))
      DO 99 I=1,K
      IF(I.EQ.J)GO TO 99
      XMP1(J)=XMP1(J)+SQRT((XNF1(I)/(XNS1+1))*(XNF2(I)/XNS2))
      XMP2(J)=XMP2(J)+SQRT((XNF1(I)/XNS1)*(XNF2(I)/(XNS2+1)))
   99 CONTINUE
  100 CONTINUE
      DO 110 I=1,K
  110 TLR(I)=ALOG(XMP1(I)/XMP2(I))
      RETURN
      END
C COMPUTES MARTIN AND BRADLEY ORTHOGONAL POLYNOMIALS
      SUBROUTINE POLY(NV,NC,N1,N2,NN1,NN2)
      DIMENSION N1(64),N2(64),XN1(64),XN2(64),FN1(64),FN2(64),
     1F(64),Y1(64),Y2(64),C1(7),C2(7),XD(7,7),
     2A1(7),A2(7),YADJ1(64),YADJ2(64),AA1(7),AA2(7),AAS1(22),
     3YAD1(64),YAD2(64),YY1(64),YY2(64),YZ1(64),YZ2(64),
     4W1(64),W2(64),W(64),YYY1(64),YYY2(64),AR1(7),AR2(7),
     7AFF1(7),AAF2(7),YYF1(64),YYF2(64),YYP1(64),YYP2(64),
     8YYS1(64),YYS2(64),AAS2(22),AF1(22),AF2(22),YFI1(64),YFI2(64)
      INTEGER D(7,7),Q(64,7),Z(64)
      WRITE(6,1120)N1
      WRITE(6,1120)N2
 1120 FORMAT(//,(1X,20I3))
      XNN1=NN1
      XNN2=NN2
      NR=NV+1
      W1=0.5
      W2=0.5
      XNN=XNN1+XNN2
      P1=XNN1/XNN
      P2=XNN2/XNN
      DO 10 I=1,NC
      XN1(I)=N1(I)
      XN2(I)=N2(I)
   10 CONTINUE
      DO 17 I=1,NC
      IF(XN1(I))11,11,12
   11 FN1(I)=0.0
      GO TO 14
   12 FN1(I)=XN1(I)/XNN1
   14 IF(XN2(I))15,15,16
   15 FN2(I)=0.0
      GO TO 17
   16 FN2(I)=XN2(I)/XNN2
   17 CONTINUE
      DO 25 I=1,NC
      F(I)=(W1*FN1(I))+(W2*FN2(I))
   25 CONTINUE
      DO 26 I=1,NC
      WRITE(6,101)FN1(I),FN2(I),F(I)
   26 CONTINUE
  101 FORMAT(/,1X,3F10.5)
      DO 31 I=1,NC
```

```
      IF(F(I))30,30,31
   30 Y1(I)=0.0
      Y2(I)=0.0
      GO TO 32
   31 Y1(I)=(FN1(I)-F(I))/F(I)
      Y2(I)=(FN2(I)-F(I))/F(I)
   32 WRITE(6,102)Y1(I),Y2(I)
   33 CONTINUE
  102 FORMAT(1H ,'Y1=',F10.5,'Y2=',F10.5)
      DO 40 I=1,NC
      IF(F(I))37,37,38
   37 Z(I)=0.0
      GO TO 40
   38 Z(I)=1.0
   40 CONTINUE
      DO 41 J=1,NC
   41 W(J,1)=1
      II=1
      DO 50 I=1,2
      DO 50 J=1,2
      DO 50 K=1,2
      DO 50 L=1,2
      DO 50 M=1,2
      DO 50 N=1,2
      W(II,2)=I
      W(II,3)=J
      W(II,4)=K
      W(II,5)=L
      W(II,6)=M
      W(II,7)=N
      II=II+1
   50 CONTINUE
  104 FORMAT(1H ,2X7(I2))
      DO 55 I=1,NC
      DO 55 J=1,NP
      IF(W(I,J).EQ.1)GO TO 51
      W(I,J)=-1
      GO TO 55
   51 W(I,J)=1
   55 CONTINUE
C FULL MODEL ESTIMATION
      WRITE(6,109)
  109 FORMAT(/,20X,'FULL MODEL ESTIMATION')
      DO 95 I=1,NP
      AA1(I)=0.0
      AA2(I)=0.0
      AAF1(I)=0.0
      AAF2(I)=0.0
   95 CONTINUE
      DO 8998 I=1,22
      AF1(I)=0.0
      AF2(I)=0.0
      AAS1(I)=0.0
 8998 AAS2(I)=0.0
      WRITE(6,9000)
 9000 FORMAT(1H ,8X,'AF1',7X,'AF2',6X,'AAF1',6X,'AAF2',
     16X,'ASS1',6X,'ASS2')
      DO 2500 J=1,NC
      YYY1(J)=0.0
      YYY2(J)=0.0
      YYF1(J)=0.0
      YYF2(J)=0.0
      YYM1(J)=0.0
      YYM2(J)=0.0
      YYS1(J)=0.0
 2500 YYS2(J)=0.0
```

```
      DO 960 I=1,NH
      DO 96 J=1,NC
      AA1(I)=AA1(I)+(Q(J,I)*Y1(J))
      AA2(I)=AA2(I)+(Q(J,I)*Y2(J))
   96 CONTINUE
      AF1(I)=AA1(I)/64.0
      AF2(I)=AA2(I)/64.0
      AAF1(I)=AA1(I)/2.0
      AAF2(I)=AA2(I)/2.0
      AAS1(I)=AA1(I)/4.0
      AAS2(I)=AA2(I)/4.0
  960 CONTINUE
      DO 970 I=1,NH
      WRITE(6,97)AF1(I),AF2(I),AAF1(I),AAF2(I),AAS1(I),AAS2(I)
  970 CONTINUE
   97 FORMAT(1H ,1X,6F10.5)
      H1=0.0
      H2=0.0
      DO 6002 J=1,NC
      H1=H1+(F(J)*Y1(J))
      H2=H2+(F(J)*Y2(J))
 6002 CONTINUE
      WRITE(6,6001)H1,H2
      CALL RISK(NC,Y1,Y2,F,P1,P2,RM)
      WRITE(6,110)RM
  110 FORMAT(1H ,5X,'RISK-FULL MODEL=',F10.5)
      DO 2501 J=1,NC
      DO 2501 I=1,NH
      YYY1(J)=YYY1(J)+(AF1(I)*Q(J,I))
      YYY2(J)=YYY2(J)+(AF2(I)*Q(J,I))
      YYF1(J)=YYF1(J)+(AAF1(I)*Q(J,I))
      YYF2(J)=YYF2(J)+(AAF2(I)*Q(J,I))
      YYM1(J)=YYM1(J)+(AAS1(I)*Q(J,I))
      YYM2(J)=YYM2(J)+(AAS2(I)*Q(J,I))
 2501 CONTINUE
      WRITE(6,9010)
 9010 FORMAT(1H ,9X,'YYY1',8X,'YYY2',8X,'YYF1',8X,'YYF2',
     18X,'YYM1',8X,'YYM2')
      DO 2502 J=1,NC
      WRITE(6,2503)YYY1(J),YYY2(J),YYF1(J),YYF2(J),YYM1(J),YYM2(J)
 2502 CONTINUE
 2503 FORMAT(1H ,1X,6F12.5)
      DO 7500 J=1,NC
      IF(YYY1(J).GT.-1)GO TO 7506
      YYY1(J)=-.99999
 7506 IF(YYY2(J).GT.-1)GO TO 7500
      YYY2(J)=-.99999
 7500 CONTINUE
      H1=0.0
      H2=0.0
      DO 6003 J=1,NC
      H1=H1+(F(J)*YYY1(J))
      H2=H2+(F(J)*YYY2(J))
 6003 CONTINUE
      WRITE(6,6001)H1,H2
      CALL RISK(NC,YYY1,YYY2,F,P1,P2,RM)
      WRITE(6,9009)RM
 9009 FORMAT(1H ,5X,'RISK-FULL MODEL-MATH=',F10.5)
 4093 FORMAT(1H ,2X,4F12.5,I5)
      II=NH
      DO 7002 I=2,NV
      KK=I+1
      DO 7002 K=KK,NH
      II=II+1
      DO 500 J=1,NC
      AF1(II)=AF1(II)+(Q(J,I)*Q(J,K)*Y1(J))
```

```
      AF2(IT)=AF2(IT)+(P(J,I)*Q(J,K)*Y2(J))
  590 CONTINUE
      AF1(IT)=AF1(IT)/64.0
      AF2(IT)=AF2(IT)/64.0
 7000 CONTINUE
 7001 FORMAT(1H ,5X,2F10.5,2I5,2F14.5)
      DO 501 I=1,NN
      WRITE(6,503)AF1(I),AF2(I)
  501 CONTINUE
      J=NN+1
      DO 502 K=J,22
      WRITE(6,503)AF1(K),AF2(K)
  502 CONTINUE
  503 FORMAT(1H ,2X,F10.5,2X,F10.5)
 7002 FORMAT(1H ,2X,4F12.5,I4)
 7003 FORMAT(1H ,2X,2F12.5)
 7004 FORMAT(1H ,10X,I4)
      DO 7005 J=1,NC
      YFI1(J)=YYY1(J)
 7005 YFI2(J)=YYY2(J)
      DO 504 J=1,NC
      II=NB+1
      DO 505 I=2,NV
      LL=I+1
      DO 506 K=LL,NP
      YFI1(J)=YFI1(J)+(AF1(II)*P(J,I)*P(J,K))
      YFI2(J)=YFI2(J)+(AF2(II)*P(J,I)*P(J,K))
      II=II+1
  506 CONTINUE
  505 CONTINUE
      II=0
  504 CONTINUE
 7006 FORMAT(1H ,21X,2I5)
 7007 FORMAT(1H ,2X,I4,4F12.5,2I8)
 7008 FORMAT(1H ,2X,I4,2F12.5)
      DO 7009 J=1,NC
      IF(YFI1(J).GT.-1)GO TO 7010
      YFI1(J)=-.99999
 7010 IF(YFI2(J).GT.-1)GO TO 7009
      YFI2(J)=-.99999
 7009 CONTINUE
      DO 507 J=1,NC
      WRITE(6,508)YFI1(J),YFI2(J)
  507 CONTINUE
  508 FORMAT(1H ,5X,'YFI1=',F10.5,5X,'YFI2=',F10.5)
      H1=0.0
      H2=0.0
      DO 7011 J=1,NC
      H1=H1+(F(J)*YFI1(J))
      H2=H2+(F(J)*YFI2(J))
 7011 CONTINUE
      WRITE(6,6001)H1,H2
      CALL RISK(NC,YFI1,YFI2,F,P1,P2,RH)
      WRITE(6,509)RH
  509 FORMAT(1H ,5X,'RISK-FULL WITH INTERACTIONS=',F10.5)
C FULL MODEL WITH INTERACTIONS
      WRITE(6,111)
  111 FORMAT(/,20X,'FIRST-ORDER MODEL')
      DO 9011 J=1,NC
      IF(YYF1(J).GT.-1)GO TO 9012
      YYF1(J)=-.99999
 9012 IF(YYF2(J).GT.-1)GO TO 9011
      YYF2(J)=-.99999
 9011 CONTINUE
      H1=0.0
      H2=0.0
```

```
      DO 8009 J=1,NC
      H1=H1+(F(J)*YYF1(J))
      H2=H2+(F(J)*YYF2(J))
 8009 CONTINUE
      WRITE(6,8001)H1,H2
      CALL RISK(NC,YYF1,YYF2,F,P1,P2,PR)
      WRITE(6,9013)PR
 9013 FORMAT(1H ,2X,'RISK=FIRST-ORDER MODEL=MAIN',F10.5)
      II=NV
      DO 1135 I=2,NV
      KK=I+1
      DO 1135 K=KK,NV
      II=II+1
      DO 135 J=1,NC
      AAS1(II)=AAS1(II)+(Q(J,I)*Q(J,K)*Y1(J))
      AAS2(II)=AAS2(II)+(Q(J,I)*Q(J,K)*Y2(J))
  135 CONTINUE
      AAS1(II)=AAS1(II)/4.0
      AAS2(II)=AAS2(II)/4.0
 1135 CONTINUE
 1500 FORMAT(1H ,5X,2F10.5,2I5,2F10.5)
      DO 136 I=1,NN
      WRITE(6,112)AAS1(I),AAS2(I)
  136 CONTINUE
      JJ=NR+1
      DO 360 K=J,22
      WRITE(6,112)AAS1(K),AAS2(K)
  360 CONTINUE
  112 FORMAT(1H ,2X,F10.5,2X,F10.5)
 1180 FORMAT(1H ,2X,4F12.5,I4)
 4500 FORMAT(1H ,2X,2F12.5)
 1190 FORMAT(1H ,10X,I4)
      DO 5556 J=1,NC
      YYS1(J)=YYM1(J)
 5556 YYS2(J)=YYM2(J)
      DO 140 J=1,NC
      II=NV+1
      DO 139 I=2,NV
      LL=I+1
      DO 138 K=LL,NV
      YYS1(J)=YYS1(J)+(AAS1(II)*Q(J,I)*Q(J,K))
      YYS2(J)=YYS2(J)+(AAS2(II)*Q(J,I)*Q(J,K))
      II=II+1
  138 CONTINUE
  139 CONTINUE
      II=0
  140 CONTINUE
 2000 FORMAT(1H ,21X,2I5)
 2001 FORMAT(1H ,2X,I4,4F12.5,2I6)
 2002 FORMAT(1H ,2X,I4,2F12.5)
      DO 7800 J=1,NC
      IF(YYS1(J).GT.-1)GO TO 7801
      YYS1(J)=-.99999
 7801 IF(YYS2(J).GT.-1)GO TO 7800
      YYS2(J)=-.99999
 7800 CONTINUE
      DO 141 J=1,NC
      WRITE(6,113)YYS1(J),YYS2(J)
  141 CONTINUE
  113 FORMAT(1H ,5X,'YYS1=',F10.5,5X,'YYS2=',F10.5)
      H1=0.0
      H2=0.0
      DO 6004 J=1,NC
      H1=H1+(F(J)*YYS1(J))
      H2=H2+(F(J)*YYS2(J))
 6004 CONTINUE
```

```
      WRITE(6,801)M1,M2
      CALL RISK(NC,YYS1,YYS2,F,P1,P2,RM)
      WRITE(6,110)RM
  110 FORMAT(1H ,5X,*RISK=SE(FUN WITH INTERACTION)=*,F10.5)
      MM=0
      DO 300 I=1,NC
      IF(F(I))320,320,300
  320 MM=MM+1
  300 CONTINUE
      IF(MM.EQ.0)RETURN
C REDUCED MODEL ESTIMATION
      WRITE(6,105)
  105 FORMAT(/,20X,*REDUCED MODEL ESTIMATION*)
      DO 60 I=1,NR
      C1(I)=0.0
      C2(I)=0.0
   60 CONTINUE
      DO 61 I=1,NR
      DO 61 J=1,NR
      D(I,J)=0.0
   61 CONTINUE
      KS=0
      DO 65 I=1,NR
      DO 65 J=1,NC
      C1(I)=C1(I)+(Z(J)*Q(J,I)*Y1(J))
      C2(I)=C2(I)+(Z(J)*Q(J,I)*Y2(J))
   65 CONTINUE
      DO 66 I=1,NR
      WRITE(6,150)C1(I),C2(I)
   66 CONTINUE
  150 FORMAT(1H ,5X,*C1=*,F10.5,5X,*C2=*,F10.5)
      DO 75 I=1,NR
      DO 75 J=1,NR
      DO 75 K=1,NC
      D(I,J)=D(I,J)+(Z(K)*Q(K,I)*Q(K,J))
   75 CONTINUE
 1130 FORMAT(1H ,61X,4I15)
  106 FORMAT(1H ,7X7(I3))
      DO 997 I=1,NR
      DO 997 J=1,NR
  997 XD(I,J)=D(I,J)
      CALL SIMQ(XD,C1,NR,KS)
      DO 85 I=1,NR
      A1(I)=C1(I)
      WRITE(6,107)A1(I)
   85 CONTINUE
  107 FORMAT(1H ,5X,F10.5)
      DO 998 I=1,NR
      DO 998 J=1,NR
  998 XD(I,J)=D(I,J)
      CALL SIMQ(XD,C2,NR,KS)
      DO 86 I=1,NR
      A2(I)=C2(I)
      WRITE(6,107)A2(I)
   86 CONTINUE
      DO 89 I=1,NC
      YADJ1(I)=0.0
   89 YADJ2(I)=0.0
      DO 90 J=1,NC
      DO 90 I=1,NR
      IF(F(J))8000,8000,8001
 8000 YADJ1(J)=YADJ1(J)+(A1(I)*Q(J,I))
      YADJ2(J)=YADJ2(J)+(A2(I)*Q(J,I))
      GO TO 90
 8001 YADJ1(J)=Y1(J)
      YADJ2(J)=Y2(J)
```

```
   90 CONTINUE
      DO 8002 J=1,NC
      IF(YADJ1(J).GT.-1)GO TO 8004
      YADJ1(J)=-.99999
 8004 IF(YADJ2(J).GT.-1)GO TO 8002
      YADJ2(J)=-.99999
 8002 CONTINUE
      DO 2000 J=1,NC
      WRITE(6,2002)YADJ1(J),YADJ2(J)
 2000 CONTINUE
 2002 FORMAT(1H ,2X,'YADJ1=',F12.5,2X,'YADJ2=',F12.5)
      H1=0.0
      H2=0.0
      DO 6000 J=1,NC
      H1=H1+(F(J)*YADJ1(J))
      H2=H2+(F(J)*YADJ2(J))
 6000 CONTINUE
      WRITE(6,6001)H1,H2
 6001 FORMAT(1H ,2X,'H1=',F15.7,2X,'H2=',F15.7)
      CALL RISK(NC,YADJ1,YADJ2,F,P1,P2,RM)
      WRITE(6,100)RM
  100 FORMAT(1H ,5X,'RISK-REDUCED MODEL=',F10.5)
      RETURN
      END
C SUBROUTINE FOR RISK ESTIMATIONS
      SUBROUTINE RISK(NC,YY1,YY2,F,P1,P2,RM)
      DIMENSION YY1(64),YY2(64),YZ1(64),YZ2(64),F(64),
     1R1(64),R2(64),R(64)
      DO 110 I=1,NC
      YZ1(I)=1.0+YY1(I)
      YZ2(I)=1.0+YY2(I)
  110 CONTINUE
  160 FORMAT(1H ,4X,4F12.5)
      RM=0.0
      DO 120 I=1,NC
      R1(I)=P1*YZ1(I)
      R2(I)=P2*YZ2(I)
  120 CONTINUE
      DO 140 I=1,NC
      IF(R1(I)-R2(I))131,131,132
  131 R(I)=R1(I)
      GO TO 140
  132 R(I)=R2(I)
  140 CONTINUE
      DO 150 I=1,NC
      RM=RM+(F(I)*R(I))
  150 CONTINUE
      RETURN
      END
/*
//GO.SYSIN DD *
CCCCCCCCCCCCC  SAMPLE    INPUT   CCCCCCCCCCCCCCCCCCCCCCCCC
    6  100  200  202 5023     1
  40   40   40   40   40   40   52   52   52   52   52   52
100
   100
      102
         100
            100
               100
100
   100
      100
         102
            100
               100
```

```
8911.       6  100   200   200  0376      1
8921.    40    40    40    40    40    40    52    52    52    52    52    52
8931.   100 10 10 10 10 10
8941.    10100 10 10 10 10
8951.    10 10100 10 10 10
8961.    10 10 10100 10 10
8971.    10 10 10 10100 10
8981.    10 10 10 10 10100
8991.   100 10 10 10 10 10
9001.    10100 10 10 10 10
9011.    10 10100 10 10 10
9021.    10 10 10100 10 10
9031.    10 10 10 10100 10
9041.    10 10 10 10 10100
9051.       6  100   200   200  5392      1
9061.    40    40    40    40    40    40    52    52    52    52    52    52
9071.   100 20 20 20 20 20
9081.    20100 20 20 20 20
9091.    20 20100 20 20 20
9101.    20 20 20100 20 20
9111.    20 20 20 20100 20
9121.    20 20 20 20 20100
9131.   100 20 20 20 20 20
9141.    20100 20 20 20 20
9151.    20 20100 20 20 20
9161.    20 20 20100 20 20
9171.    20 20 20 20100 20
9181.    20 20 20 20 20100
9191.       6  100   200   200  4021      1
9201.    40    40    40    40    40    40    52    52    52    52    52    52
9211.   100 30 30 30 30 30
9221.    30100 30 30 30 30
9231.    30 30100 30 30 30
9241.    30 30 30100 30 30
9251.    30 30 30 30100 30
9261.    30 30 30 30 30100
9271.   100 30 30 30 30 30
9281.    30100 30 30 30 30
9291.    30 30100 30 30 30
9301.    30 30 30100 30 30
9311.    30 30 30 30100 30
9321.    30 30 30 30 30100
9331.       6  100   200   200  2359      1
9341.    40    40    40    40    40    40    52    52    52    52    52    52
9351.   100 33 33 33 33 33
9361.    33100 33 33 33 33
9371.    33 33100 33 33 33
9381.    33 33 33100 33 33
9391.    33 33 33 33100 33
9401.    33 33 33 33 33100
9411.   100 33 33 33 33 33
9421.    33100 33 33 33 33
9431.    33 33100 33 33 33
9441.    33 33 33100 33 33
9451.    33 33 33 33100 33
9461.    33 33 33 33 33100
9471.       6  100   400   400  5423      1
9481.    40    40    40    40    40    40    52    52    52    52    52    52
9491.   100
9501.      100
9511.         100
9521.            100
9531.               100
9541.                  100
9551.   100
9561.      100
```

```
 9571.           100
 9581.              1.0
 9591.                 130
 9601.                    100
 9611.        6  100  400  400 0376     1
 9621.     40   40   40   40   40   40   52   52   52   52   52   52
 9631.    100 10 10 10 10 10
 9641.     10100 10 10 10 10
 9651.     10 10100 10 10 10
 9661.     10 10 10100 10 10
 9671.     10 10 10 10100 10
 9681.     10 10 10 10 10100
 9691.    100 10 10 10 10 10
 9701.     10100 10 10 10 10
 9711.     10 10100 10 10 10
 9721.     10 10 10100 10 10
 9731.     10 10 10 10100 10
 9741.     10 10 10 10 10100
 9751.        6  100  400  400 5302     1
 9761.     40   40   40   40   40   40   52   52   52   52   52   52
 9771.    100 20 20 20 20 20
 9781.     20100 20 20 20 20
 9791.     20 20100 20 20 20
 9801.     20 20 20100 20 20
 9811.     20 20 20 20100 20
 9821.     20 20 20 20 20100
 9831.    100 20 20 20 20 20
 9841.     20100 20 20 20 20
 9851.     20 20100 20 20 20
 9861.     20 20 20100 20 20
 9871.     20 20 20 20100 20
 9881.     20 20 20 20 20100
 9891.        6  100  400  400 0821     1
 9901.     40   40   40   40   40   40   52   52   52   52   52   52
 9911.    100 30 30 30 30 30
 9921.     30100 30 30 30 30
 9931.     30 30100 30 30 30
 9941.     30 30 30100 30 30
 9951.     30 30 30 30100 30
 9961.     30 30 30 30 30100
 9971.    100 30 30 30 30 30
 9981.     30100 30 30 30 30
 9991.     30 30100 30 30 30
10001.     30 30 30100 30 30
10011.     30 30 30 30100 30
10021.     30 30 30 30 30100
10031.        6  100  400  400 2359     1
10041.     40   40   40   40   40   40   52   52   52   52   52   52
10051.    100 33 33 33 33 33
10061.     33100 33 33 33 33
10071.     33 33100 33 33 33
10081.     33 33 33100 33 33
10091.     33 33 33 33100 33
10101.     33 33 33 33 33100
10111.    100 33 33 33 33 33
10121.     33100 33 33 33 33
10131.     33 33100 33 33 33
10141.     33 33 33100 33 33
10151.     33 33 33 33100 33
10161.     33 33 33 33 33100
10171.       -1
10181.    /*
10191.    //
```

REFERENCES

REFERENCES

Bahadur, R.R. "A Representation of the Joint Distribution of Responses to Dichotomous Items," in "Studies in Item Analysis and Prediction." Stanford University Press, 1961, pp. 169-76.

Banks, S. "The Relationship Between Preference and Purchase of Brands," Journal of Marketing, Vol. 15 (October 1951), pp. 145-57.

Benson, P.H. "How Many Scales and How Many Categories Shall We Use in Consumer Research--A Comment," Journal of Marketing, Vol. 35 (October 1971), pp. 59-61.

Bonfield, E.H. "Attitude, Social Influence, Personal Norm, and Intention Interactions as Related to Brand Purchase Behavior," Journal of Marketing Research, Vol. 11 (November 1974), pp. 379-89.

Boyd, H.W. and R. Westfall. Marketing Research (Homewood, Illinois: Richard D. Irwin, 1972).

Brody, R.P. and S.M. Cunningham. "Personality Variables and the Consumer Decision Process," Journal of Marketing Research, Vol. 5 (February 1968), pp. 50-57.

Brown, L.O. and L.L. Beik. Marketing Research and Analysis, 4th ed. (New York: The Ronald Press Company, 1969).

Cascoullas, T. "Estimation of a Multivariate Density," Annual Inst. Stat. Math., Vol. 18, No. 2 (1966), pp. 179-89.

Claycamp, H.R. "Characteristics of Owners of Thrift Deposits in Commercial Banks and Savings and Loan Associations," Journal of Marketing Research, Vol. 2 (May 1965), pp. 163-70.

Cochran, William G. and C.E. Hopkins. "Some Classification Problems with Multivariate Qualitative Data," Biometrics, Vol. 17 (March 1961), pp. 10-32.

Darden, W.R. and F.D. Reynolds. "Backward Profiling of Male Innovators," Journal of Marketing Research, Vol. 11 (February 1974), pp. 79-86.

Dash, J. Store Choice: An Investigation of the Characteristics of Consumers Who Bought Audio Equipment from a Specialty Retailer versus a Department Store. Unpublished Ph.D. dissertation, The City University of New York, 1974.

Dodge, H.R. and H.H. Summer. "Choosing Between Retail Stores," Journal of Retailing, Vol. 45 (Fall 1969), pp. 11-21.

Elashoff, I.D., R.M. Elashoff and G.E. Goldman. "On the Choice of Variables in Classification Problems with Dichotomous Variables," Biometrika, Vol. 54 (1957), pp. 668-70.

Etzel, M.J. "Using Multiple Discriminant Analysis to Segment the Consumer Credit Market," 1974 Combined Proceedings (Chicago: American Marketing Association, 1974), pp. 35-40.

Evans, F.B. "Psychological and Objective Factors in the Prediction of Brand Choice: Ford versus Chevrolet," Journal of Business, Vol. 31 (October 1959), pp. 340-69.

________ "Reply: You Still Can't Tell a Ford Owner from a Chevrolet Owner," Journal of Business, Vol. 34 (January 1961), pp. 67-73.

________ "Ford versus Chevrolet: Park Forest Revisited," Journal of Business, Vol. 40 (October 1968), pp. 445-59.

Fisher, R.A. "The Use of Measurements in Taxonomic Problems," Annuals of Eugenics, Vol. 7 (1936), pp. 176-84.

________ Statistical Methods for Research Workers. (New York: Haftner Publishing Company, 1958).

Fix, E. and J.L. Hodges. "Nonparametric Discrimination Consistency Properties," U.S. Air Force School of Aviation Medicine, Report No. 4, 1951.

Frank, R.E. "Market Segmentation Research: Findings and Implications," in F. Bass et al (eds.), Application of the Sciences to Marketing Management (New York: John Wiley and Sons, 1968).

________ and W.F. Massy. "Innovation and Brand Choice: The Folger's Invasion," in Greyser (ed.), Toward Scientific Marketing (Chicago: American Marketing Association, 1963).

Frank, R.E., W.F. Massy and H.W. Boyd. "Correlates of Grocery Product Consumption Rates," Journal of Marketing Research (May 1967), pp. 184-90.

Frank, R.E., W.F. Massy and T.W. Lodhal. "Purchasing Behavior and Personal Attributes," Journal of Advertising Research, Vol. 9 (December 1969), pp. 15-24.

Frank, R.E., W.F. Massy and D.G. Morrison. "The Determinants of Innovative Behavior with Respect to A Branded Frequently Purchased Food Product," in L.G. Smith (ed.), Reflections on Progress in Marketing (Chicago: American Marketing Association, 1965).

__________ "Bias in Multiple Discriminant Analysis," Journal of Marketing Research, Vol. 2 (August 1965), pp. 250-58.

Frank, R.E., W.F. Massy and Y. Wind. Market Segmentation (Englewood Cliffs, New Jersey: Prentice-Hall, 1972).

Fukunga, K. and D. Kessel. "Error Evaluation and Model Validation in Statistical Pattern Recognition," School of Electrical Engineering, Purdue University, Tech Rep. TR-EE, 1972, pp. 72-23.

Gatty, Ronald. "Multivariate Analysis for Marketing Research: An Evaluation," Journal of the Royal Statistical Society, Series C, Vol. 15 (November 1960), pp. 157-72.

Gessaman, M.P. "A Consistent Nonparametric Multivariate Density Estimator Based on Statistically Equivalent Blocks," Annual of Mathematical Statistics, Vol. 41 (1970), pp. 1344-46.

__________ and P.H. Gessaman. "A Comparison of Some Multivariate Discrimination Procedures," Journal of the American Statistical Association, Vol. 67 (1972), pp. 468-72.

Gilbert, Ethel S. "On Discrimination Using Qualitative Variables," Journal of American Statistical Association, Vol. 67 (1972), pp. 116-22.

Glick, N. "Sample-based, Multinomial Classification," Biometrics, Vol. 29 (1973), pp. 241-56.

Goldstein, M. "Comparison of Some Density Estimate Classification Procedures," Journal of the American Statistical Association, Vol. 70 (September 1975), pp. 666-69.

________ and M. Rabinowitz. "Selection of Variates for the Two-Group Multinomial Classification Problem," Journal of American Statistical Association, Vol. 70 (December 1975), pp. 776-81.

Goldstein, M., E. Wolf and W. R. Dillon. "On a Test of Independence for Contingency Tables," Communications in Statistics (February 1976)

Green, P.E. and V.R. Rao. "Rating Scales and Information Recovery--How Many Scales and Response Categories to Use?" Journal of Marketing, Vol. 34 (July 1970), pp. 33-9.

________ and D.S. Tull. Research for Marketing Decisions (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975).

Hills, M. "Allocation Rules and Their Error Rates," Journal of the Royal Statistical Society, Series B (1966), p. 1.

________ "Discrimination and Allocation with Discrete Data," Journal of the Royal Statistical Society, Series C (1967), pp. 237-50.

Ito, R. "Differential Attitudes of New Car Buyers," Journal of Advertising Research, Vol. 7 (March 1967), pp. 38-42.

Jacoby, J. and M.S. Matell. "Three Point Likert Scales Are Good Enough," Journal of Marketing Research, Vol. 8 (November 1971), pp. 495-500.

Johnston, J. Econometric Methods, 2nd ed. (New York: McGraw-Hill Book Company, 1972).

Keuhn, A.A. "Demonstration of a Relationship Between Psychological Factors and Brand Choice," Journal of Business, Vol. 36 (April 1963), pp. 237-41.

King, A. and D.W. King. "A Mathematical Model to Isolate the Factors which Determine Market Shares," in R.M. Kaplan (ed.), The Marketing Concept in Action (Chicago: American Marketing Association, 1964).

King, C.W. "The Innovator in the Fashion Adoption Process," in L.G. Smith (ed.), Reflections on Progress in Marketing (Chicago: American Marketing Association, 1964).

King, W.R. "Marketing Expansion - A Statistical Analysis," Management Science, Vol. 9 (July 1963), pp. 563-73.

__________ "Evaluation in Marketing Systems," Management Science, Vol. 10 (July 1964), pp. 659-66.

__________ "Early Prediction of New Product Success," Journal of Advertising Research, Vol. 6, No. 2, (June 1966), pp. 8-13.

__________ "On Methods: Structural Analysis and Descriptive Discriminant Function," Journal of Advertising Research, Vol. 7, No. 2 (June 1967), pp. 39-43.

Krzanowski, W.J. "Discrimination and Classification Using Both Binary and Continuous Variables," Journal of American Statistical Association, Vol. 70 (December 1975), pp. 782-90.

Lachenbruch, P.A. Discriminant Analysis (New York: Hafner Press, 1975).

Marcus, A.S. Advertising Age (January 18, 1960), p. 92.

__________ "Obtaining Group Measures from Personality Test Scores," Psychological Reports, Vol. 17 (1965), pp. 523-31.

Martilla, J.A. and D.W. Carvey. "Four Subtle Sins in Marketing Research," Journal of Marketing, Vol. 39 (January 1975), pp. 8-15.

Martin, D.C. and R.A. Bradley. "Probability Models Estimation and Classification for Multivariate Dichotomous Populations," Biometrics, Vol. 28 (March 1972), pp. 203-21.

Martin, W.S. "The Effects of Scaling on the Correlation Coefficient: A Test of Validity," Journal of Marketing Research, Vol. 10 (August 1973), pp. 316-18.

Martineau, P. Advertising Age (December 21, 1959), p. 76.

Massy, W.R. "On Methods: Discriminant Analysis of Audience Characteristics," Journal of Advertising Research, Vol. 5 (March 1965), pp. 39-45.

Matusita, K. "On the Theory of Statistical Decision Functions," Annual Inst. Stat. Math., Vol. 3 (1951), pp. 17-35.

__________ and H. Akaike. "Note on the Decision Problem," Annual Inst. Stat. Math., Vol. 4 (1952), pp. 11-14.

__________ "Decision Rule, Based on Distance, for the Classification Problem," Annual Inst. Stat. Math., Vol. 8 (1956), pp. 67-77.

________ and M. Motoo. "On the Fundamental Theorem for the Decision Rule Based on Distance," Annual Inst. Stat. Math., Vol. 7 (1956), pp. 137-42.

Montgomery, D.B. "New Product Diffusion - An Analysis of Supermarket Buyer Decision," Journal of Marketing Research, Vol. 12 (August 1975), pp. 255-65.

Moore, D. Evaluation of Five Discrimination Procedures for Binary Variables. Unpublished Ph.D. dissertation, University of California, Berkeley, 1970.

________ "Evaluation of Five Discrimination Procedures for Binary Variables," Journal of the American Statistical Association, Vol. 68 (June 1973), pp. 399-404.

Morrison, D.G. "On the Interpretation of Discriminant Analysis," Journal of Marketing Research, Vol. 6 (May 1969), pp. 156-63.

Muczyk, J.F., T.H. Mattheiss and M. Gable. "Predicting Success of Store Managers," Journal of Retailing, Vol. 50 (Summer 1974), pp. 43-50.

Myers, J.H. and G. Warner. "Semantic Properties of Selected Evaluation Adjectives," Journal of Marketing Research, Vol. 5 (November 1968), pp. 409-12.

Parzen, E. "An Estimation of a Probability Density Mode," Annual of Mathematical Statistics, Vol. 3 (1962), pp. 1065-76.

Perry, M. "Discriminant Analysis of Relations Between Consumer's Attitude Behavior and Intention," Journal of Advertising Research, Vol. 9, No. 2 (1969), pp. 34-9.

Pessemier, E.A., P.C. Burger and D.J. Tigert. "Can New Product Buyers Be Identified,""Journal of Marketing Research, Vol. 4 (November 1967), pp. 349-54.

Philpot, J.W., R.C. Reizenstein and D.J. Sweeney. "Identifying Determinants of Store Patronage Using Factor Analysis." A paper presented at the Third Annual Conference of the Association for Consumer Research, Chicago (November 1972).

Press, James C. Applied Multivariate Analysis. (New York: Holt, Reinhart and Winston, 1972).

Revo, L.T. "On Classifying with Certain Types of Ordered Qualitative Variates: An Evaluation of Several Procedures," North Carolina Institute Statistical Mimeo Service, No. 708, 1970.

Robertson, T.S. and J.N. Kennedy. "Prediction of Consumer Innovators: Application of Multiple Discriminant Analysis," Journal of Marketing Research, Vol. 5 (February 1968), pp. 64-9.

Seth, J.N. An Investigation into the Mediating Effects of Socio-Psychological Variables Between Advertising Stimulus and Brand Loyalty. Unpublished Ph.D. dissertation, Columbia University, 1958.

Sheth, J.N. "Multivariate Analysis in Marketing," Journal of Advertising Research, Vol. 10, No. 1 (February 1970), pp. 29-37.

Smith, W.R. "Product Differentiation and Marketing Segmentation as Alternative Marketing Strategies," Journal of Marketing, Vol. 21 (July 1956), pp. 3-8.

Steiner, G. "Notes on Franklin B. Evans' Psychological and Objective Factors in the Prediction of Brand Choice," Journal of Business, Vol. 34 (January 1961), pp. 57-60.

Stevens, S.S. "Measurement, Statistics and the Schematic View," Science, Vol. 161 (August 1968), pp. 849-56.

Uhl, K., R. Andrus and L. Paulsen. "How Are Laggards Different? An Empirical Inquiry," Journal of Marketing Research, Vol. 7 (February 1970), pp. 51-4.

Welch, B.L. "Note on Discriminant Functions," Biometrika, Vol. 2 (1939), pp. 218-20.

Wells, W.D. and D.J. Tigert. "Activities, Interests and Opinions," Journal of Advertising Research, Vol. 11 (December 1971), pp. 27-35.

Westfall, R. "Psychological Factors in Predicting Product Choice," Journal of Marketing, Vol. 36 (April 1962), pp. 34-40.

Winich, C. "The Relationship Among Personality Needs, Objective Factors and Brand Choice: A Reexamination," Journal of Business, Vol. 34 (January 1961), pp. 61-6.

Yankelovich, D. "New Criteria for Market Segmentation," Harvard Business Review, Vol. 42 (March-April 1964), pp. 83-90.